Bayesian Stochastic Mortality Modelling under Serially Correlated Local Effects

George Mavros

Submitted for the degree of
Doctor of Philosophy

Heriot-Watt University

Department of Actuarial Mathematics and Statistics,
School of Mathematical and Computer Sciences.

January 12, 2015

Abstract

The vast majority of stochastic mortality models in the academic literature are intended to explain the dynamics underpinning the process by a combination of age, period and cohort effects. In principle, the more such effects are included in a stochastic mortality model, the better is the in-sample fit to the data. Estimates of those parameters are most usually obtained under some distributional assumption about the occurrence of deaths, which leads to the optimisation of a relevant objective function.

The present Thesis develops an alternative framework where the local mortality effect is appreciated, by employing a parsimonious multivariate process for modelling the latent residual effects of a simple stochastic mortality model as dependent rather than conditionally independent variables. Under the suggested extension the cells of the examined data-set are supplied with a serial dependence structure by relating the residual terms through a simple vector autoregressive model. The method is applicable for any of the popular mortality modelling structures in academia and industry, and is accommodated herein for the Lee-Carter and Cairns-Blake-Dowd models. The additional residuals model is used to compensate for factors of a mortality model that might mostly be affected by local effects within given populations. By using those two modelling bases, the importance of the number of factors for a stochastic mortality model is emphasised through the properties of the prescribed residuals model. The resultant hierarchical models are set under the Bayesian paradigm, and samples from the joint posterior distribution of the latent states and parameters are obtained by developing Markov chain Monte Carlo algorithms. Along with the imposed short-term dynamics, we also examine the impact of the joint estimation in the long-term factors of the original models. The Bayesian solution aids in recognising the different levels of uncertainty for the two naturally distinct type of dynamics across different populations. The forecasted rates, mortality improvements, and other relevant mortality dependent metrics under the developed models are compared to those produced by their benchmarks and other standard stochastic mortality models in the literature.

Acknowledgements

The author is indebted to his supervisors, Dr. George Streftaris, Prof. Andrew Cairns and Dr. Torsten Kleinow for their guidance, support and encouragement throughout the completion of this research.

Contents

Abstract ii

Acknowledgements iii

Introduction 1

1 Mortality Modelling 12

1.1 Introduction . . . 13

1.1.1 Mortality Rates . . . 13

1.1.2 Force of Mortality . . . 14

1.1.3 Mortality Laws . . . 16

1.1.4 Modelling Assumptions . . . 18

1.1.5 Mortality Metrics . . . 20

1.2 The Lee-Carter Model . . . 23

1.2.1 Description . . . 24

1.2.2 Estimation . . . 25

1.2.3 Summary and Extensions . . . 30

1.3 The Cairns-Blake-Dowd Model . . . 35

1.3.1 Description . . . 36

1.3.2 Estimation . . . 36

1.3.3 Summary and Extensions . . . 39

1.4 Comparisons and Forecasts . . . 43

1.4.1 Comparison of Mortality Models . . . 43

1.4.2 Projections of Mortality Rates . . . 47

1.4.3 Further Projections and Numerical Examples . . . 56

1.5 Summary . . . 58

2 Bayesian Modelling and MCMC methods 60

2.1 Introduction . . . 61

2.1.1 Bayesian Analysis . . . 61

2.1.2 Prior specification . . . 62

2.1.3 Hierarchical Modelling . . . 63

2.2 Markov chain Monte Carlo methods . . . 64
2.2.1 Markov Chain theory . . . 65
2.2.2 Markov Chain Monte Carlo . . . 67
2.2.3 Practical Considerations . . . 71
2.3 MCMC for State Space Models . . . 73
2.3.1 Model Specification . . . 73
2.3.2 Bayesian Estimation . . . 73
2.3.3 Convergence and summary of the posterior distribution . . . 75
2.4 Parameter Uncertainty in Mortality Modelling . . . 80
2.4.1 Bayesian CBD Model . . . 81
2.4.2 Convergence & Posterior Inference for the Bayesian CBD Model 84
2.4.3 Comparative Simulations and Forecasts . . . 89
2.5 Summary . . . 93

3 Residuals Modelling 94
3.1 Introduction . . . 95
3.2 Residual Analysis . . . 96
3.2.1 Visual Diagnostics . . . 96
3.2.2 Statistical Analysis . . . 97
3.2.3 Correlation Analysis . . . 99
3.3 The Residuals Process . . . 100
3.3.1 Model Description . . . 102
3.3.2 Empirical Analysis . . . 103
3.4 Summary . . . 107

4 Bayesian Augmented Lee-Carter Model 109
4.1 Model description . . . 110
4.2 Model fitting . . . 111
4.2.1 Implementation . . . 112
4.2.2 Posterior simulation . . . 116
4.3 Application . . . 122
4.3.1 Convergence . . . 122
4.3.2 Posterior estimates . . . 127
4.4 Summary . . . 136

5 Bayesian Augmented CBD Model 138
5.1 Model description . . . 139
5.2 Model fitting . . . 140
5.2.1 Implementation . . . 140
5.2.2 Posterior simulation . . . 142

5.3 Application . . . 146
5.3.1 Convergence . . . 146
5.3.2 Posterior estimates . . . 150
5.4 Summary . . . 156

6 Comparison of proposed models **157**
6.1 Goodness-of-fit . . . 158
6.1.1 Fitted rates . . . 158
6.1.2 Standardised residuals . . . 161
6.2 Model dynamics . . . 163
6.2.1 Long-term dynamics . . . 163
6.2.2 Local dynamics . . . 166
6.3 Deviance information criterion . . . 169
6.4 Forecasted rates . . . 171
6.5 Mortality metrics . . . 178
6.6 Robustness . . . 190
6.7 Summary . . . 197

Epilogue **199**

Appendix A: Markov Chains **203**

Appendix B: Statistical Tests **204**

Bibliography **208**

List of Figures

1 Period life expectancy at birth of EW males and females for the period 1841-2011 as estimated by the Human Mortality Database. 3

2 Crude log-mortality surface and age-specific crude log-death rates for the EW data on the top and bottom panels, respectively. 5

1.1 Series of parameter estimates, $\beta_x^{(1)}$, $\beta_x^{(2)}$ and $\kappa_t^{(2)}$, of the LC model, EW and Canada data. 29

1.2 Series of parameter estimates, $\beta_x^{(1)}$, $\kappa_t^{(2)}$ and $\gamma_c^{(3)}$, of the APC model, EW and Canada data. 35

1.3 Series of parameter estimates, $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$, of the CBD model, EW and Canada data. 39

1.4 Series of parameter estimates, $\kappa_t^{(1)}$, $\kappa_t^{(2)}$, $\kappa_t^{(3)}$ and $\gamma_c^{(4)}$, of model M7, EW and Canada data. 42

1.5 Standardised residuals of deaths from the LC, APC, CBD and M7 models, EW and Canada data. Red and black squares indicate positive and negative values, respectively. 45

1.6 Estimated cohort process, $\gamma_c^{(3)}$, of the APC model for EW and Canada populations, along with fan charts of future dynamics under the assigned ARIMA specifications. 49

1.7 Residuals analysis of ARIMA(1,1,2) model for the cohort parameters, $\gamma_c^{(3)}$, of the APC model for EW and Canada populations, based on plots of estimated standardised values, ACF's and PACF's. 50

1.8 Projected central death rates on the log-scale for ages 65, 75 and 85 over a 50 years ahead horizon for the LC and APC models, EW data. 52

1.9 Projected mortality rates on the logit-scale for ages 65, 75 and 85 over a 50 years ahead horizon for the CBD and M7 models, EW data. . . 53

1.10 Projected central death rates on the log-scale for ages 65, 75 and 85 over a 50 years ahead horizon for the LC and APC models, Canada data. 54

1.11 Projected mortality rates on the logit-scale for ages 65, 75 and 85 over a 50 years ahead horizon for the CBD and M7 models, Canada data. 55

1.12 Projections of survivor indices, $\mathcal{S}(65,t)$, and associated confidence bands over the next 25 years for the LC, APC, CBD and M7 mortality models, EW and Canada data. 56

2.1 Cumulative trace-plots of means and of the associated 95% HPD intervals for parameters μ, σ_x^2 and σ_y^2 and for latent states x_6, x_{16} and x_{26} of the Gaussian State Space model.The red dashed lines indicate the original values of parameters and latent states. 76
2.2 Upper two panels: Posterior densities for parameters μ, σ_x^2 and σ_y^2 and for latent states x_6, x_{16} and x_{26} of the Gaussian State Space model. Dashed lines indicate the 95% HPD bands and the red line shows original values used for the simulation. Lower two panels: Autocorrelation functions for parameters μ, σ_x^2 and σ_y^2 and for latent states x_6, x_{16} and x_{26} of the Gaussian State Space model. 78
2.3 Contours of the cross-correlation matrix for the vector of latent states, $\boldsymbol{x}$, of the Gaussian State Space model. 79
2.4 Plot of simulated values of the vector $\boldsymbol{x}$ in dots and of the posterior means and 95% credibility intervals from the MCMC sample in solid and dashed lines, respectively. 80
2.5 Cumulative trace-plots of mean and 95% HPD intervals for parameters of the bivariate random walk, $\boldsymbol{\delta}$ and V_ζ, and of the vector of period effects, $\boldsymbol{\kappa}_{2009}$, of the B-CBD model, EW data. 86
2.6 Autocorrelation functions of the MCMC sample for the drift vector, $\boldsymbol{\delta}$, and the elements of the covariance matrix, V_ζ, of the B-CBD model, EW data. 88
2.7 Posterior means of the period effects, $\boldsymbol{\kappa}_t$, of the B-CBD model in blue dots along with the MLE's of the original CBD model under the Poisson model in red circles, EW data. 89

3.1 Contour plots of standardised residuals of the LC and the CDB models on the left and right, respectively, EW data. 97
3.2 Standardised residuals of the LC and CBD models against the years, the ages and the years of births of the cohorts, EW data. 98
3.3 Empirical one-step ahead correlation matrices of the applied autoregression for the raw residuals of the LC and CBD models, EW data. . 107
3.4 Diagonal elements of empirical covariance matrix, $\widehat{V}_Z$, of the VAR residuals model under the LC and CBD benchmarks, EW data. 108

4.1 Trace-plots of age effects, $\beta_x^{(1)}$ and $\beta_x^{(2)}$, for $x = 65, 75$ and 85, of latent state $\kappa_{2009}^{(2)}$, and of drift and variance parameters, δ and σ_κ^2, of BA-LC model, EW data. The cumulative trace-plots of δ and σ_κ^2 also include the Poisson MLE's of the original LC model in red. 124
4.2 Trace-plots of age effects, $\beta_x^{(1)}$ and $\beta_x^{(2)}$, for $x = 65, 75$ and 85, of latent state $\kappa_{2009}^{(2)}$, and of drift and variance parameters, δ and σ_κ^2, of BA-LC model, Canada data. The cumulative trace-plots of δ and σ_κ^2 also include the Poisson MLE's of the original LC model in red. 125
4.3 Trace-plots and cumulative trace-plots of autoregression parameters α_4, α_5 and α_6 of BA-LC model, EW and Canada data. 126
4.4 Posterior means and 95% credibility bands for $\beta_x^{(1)}$, $\beta_x^{(2)}$ and $\kappa_t^{(2)}$ of BA-LC model in black solid and dashed lines, respectively, for EW and Canada data. The original LC MLE's are plotted in blue and the adjusted LC estimates under the constraints of the BA-LC model are shown in red circles. 130
4.5 From top to bottom: Posterior means of residual sates, $\mathcal{R}(x,t)$ Posterior mean of matrix V_Z, scaled to represent correlations Posterior mean of matrix $V_\mathcal{R}$, scaled to represent correlations, EW and Canada data left and right panels respectively. 132
4.6 Box-plots for posteriors and posterior correlation matrices of diagonal entries of $V_\mathcal{R}$ matrix for EW and Canada data. 134
4.7 Bivariate scatterplots for posterior distributions of autoregression parameters, α_2 and α_3, and α_4, α_5 and α_6, of the VAR model for the residuals, $\mathcal{R}_t$, for EW and Canada data. 136

5.1 Cumulative trace-plots for the drift vector, $\boldsymbol{\delta}$, for the variance parameter, v, of matrix V_Z and for the autoregression coefficients α_4, α_5 and α_6 of the BA-CBD model, EW data. The red lines indicate the Poisson MLE's of the original CBD model. 148
5.2 Cumulative trace-plots for the drift vector, $\boldsymbol{\delta}$, for the variance parameter, v, of matrix V_Z and for the autoregression coefficients α_4, α_5 and α_6 of the BA-CBD model, Canada data. The red lines indicate the Poisson MLE's of the original CBD model. 149
5.3 Posterior means and 95% credibility bands for $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ of BA-CBD model in black solid and dashed lines, respectively, for EW and Canada data. The original CBD and model's M7 MLE's are plotted in red and blue, respectively. 152
5.4 Box-plots for posteriors and posterior correlation matrices of diagonal entries of $V_\mathcal{R}$ matrix for EW and Canada data. 153

5.5 Bivariate scatterplots for posterior distributions of the drift vectors, $\boldsymbol{\delta}$, and the autoregression parameters, α_2 and α_3, and α_4, α_5 and α_6, for EW and Canada data. 155

6.1 Observed log-death-rates and fitted estimates of the LC, CBD, APC, M7, BA-CBD and BA-LC models for ages 60-89 in year 2009, EW and Canada. 159
6.2 Projected mean mortality improvement rates and associated 95% confidence bands for ages 60-89 in year 2009, EW and Canada data. . . . 160
6.3 Graphs of posterior means of standardised residuals for the BA-CBD and BA-LC models for EW and Canada data, indicated in red and black squares for positive and negative values, respectively. 161
6.4 Marginal posterior densities of the long-term dynamics parameters, $\boldsymbol{\delta}$ and V_ζ, of the BA-CBD model for EW and Canada data. 164
6.5 Marginal posterior densities of the long-term dynamics parameters, δ and σ^2_κ, of the BA-LC model for EW and Canada data. 165
6.6 Marginal posterior densities for the parameters of the autoregression matrix, $\mathcal{A}$, of the BA-models for EW and Canada data. 167
6.7 Marginal posterior summaries for the covariance structures of the BA-models for EW and Canada data. 168
6.8 Forecasted mean mortality rates with their associated 95% confidence bands on the logit-scale for ages 65, 75 and 85 over a 50 years ahead horizon for the BA-CBD and BA-LC models, EW data. 172
6.9 Forecasted mean mortality rates with their associated 95% confidence bands on the logit-scale for ages 65, 75 and 85 over a 50 years ahead horizon for the BA-CBD and BA-LC models, Canada data. 173
6.10 Forecasted mean mortality rates with their associated 95% confidence bands on the logit-scale for ages 65, 75 and 85 over a 50 years ahead horizon for the CBD, M7 and BA-CBD models, EW data. 174
6.11 Forecasted mean death rates with their associated 95% confidence bands on the log-scale for ages 65, 75 and 85 over a 50 years ahead horizon for the LC, APC and BA-LC models, EW data. 175
6.12 Forecasted mean mortality rates with their associated 95% confidence bands on the logit-scale for ages 65, 75 and 85 over a 50 years ahead horizon for the CBD, M7 and BA-CBD models, Canada data. 176
6.13 Forecasted mean death rates with their associated 95% confidence bands on the log-scale for ages 65, 75 and 85 over a 50 years ahead horizon for the LC, APC and BA-LC models, Canada data. 177

6.14 Projections of survivor indices, S(65, t), and associated 95% confidence bands over the next 25 years for the BA-CBD, BA-LC, LC, APC, CBD and M7 mortality models, for the EW and Canada data-sets. 178
6.15 Density plots of unit 25 years term-annuities payable in arrears starting in 2010 under a constant 4% interest rate and cohort life-expectancies for individuals aged 65 in 2010 for the BA-CBD, BA-LC, LC, APC, CBD and M7 mortality models, for the EW and Canada data-sets. . 180
6.16 Forecasted densities of the central death rates, $m(x, 2010)$ for $x = 65, 75, 85$ under the BA-CBD, BA-LC, LC, APC, CBD and M7 mortality models, EW data. The dashed vertical line in each plot indicates the realised rate during 2010. 181
6.17 Projected mean improvement rates of $q(x, t)$ for $x = 65, 75, 85$ and associated 95% confidence bands over the period from 2009 to 2058, EW data. 183
6.18 Projected mean improvement rates of $q(x, t)$ for $x = 65, 75, 85$ and associated 95% confidence bands over the period from 2009 to 2058, Canada data. 184
6.19 Projected mean improvement rates of $q(x, t)$ for all ages, $x = 60, \ldots, 89$ for the BA-CBD, CBD and M7 models, EW data. The left and right panels show the improvements across the full projection period and during the final 20 years, respectively. 186
6.20 Projected mean improvement rates of $q(x, t)$ for all ages, $x = 60, \ldots, 89$ for the BA-LC, LC and APC models, EW data. The left and right panels show the improvements across the full projection period and during the final 20 years, respectively. 187
6.21 Projected mean improvement rates of $q(x, t)$ for all ages, $x = 60, \ldots, 89$ for the BA-CBD, CBD and M7 models, Canada data. The left and right panels show the improvements across the full projection period and during the final 20 years, respectively. 188
6.22 Projected mean improvement rates of $q(x, t)$ for all ages, $x = 60, \ldots, 89$ for the BA-LC, LC and APC models, Canada data. The left and right panels show the improvements across the full projection period and during the final 20 years, respectively. 189

6.23 Top panel: Posterior distributions of the period effects series, $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$, of the BA-CBD model for the EW data, using data 1960-2009 and 1960-1999, in black and red lines respectively. Solid lines indicate mean estimates and dashed lines show the 95% credible intervals. Middle and bottom panels: Posterior densities of the autoregression coefficients, α_4, α_5 and α_6, and of the variance parameter, v, of the BA-CBD model using the full and truncated EW data. 191
6.24 Posterior distributions of age effects series, $\beta_x^{(1)}$ and $\beta_x^{(2)}$, and of the period effects series, $\kappa_t^{(2)}$, of the BA-LC model for the EW data, using data 1960-2009 and 1960-1999, in black and red lines respectively. Solid lines indicate mean estimates and dashed lines show the 95% credible intervals. 193
6.25 Posterior densities of the autoregression coefficients, α_4, α_5 and α_6, of the BA-LC model using the full and truncated EW data, and mean posterior estimate of the covariance matrix V_Z for the truncated EW data. 194
6.26 In sample forecasts for the BA-CBD, CBD and M7 models using the truncated EW data, compared to the forecast of the BA-CBD model under the full EW data. Solid lines indicate the mean projection and the dashed lines show the 95% confidence bands. 195
6.27 In sample forecasts for the BA-LC, LC and APC models using the truncated EW data, compared to the forecast of the BA-LC model under the full EW data. Solid lines indicate the mean projection and the dashed lines show the 95% confidence bands. 196

List of Tables

1 Period life expectancies at birth and at age 65 for EW males and females as estimated by the ONS. 2

1.1 Summary statistics of estimated standardised residuals of deaths for models LC, APC, CBD and M7, EW and Canada data. 44
1.2 BIC criterion and ranking amongst models LC, APC, CBD and M7, EW and Canada data. 46
1.3 Means and standard deviations of cohort life-expectancies and of values of annual term annuities payable in arrears for a maximum of 25 years, P, for the LC, APC, CBD and M7 mortality models, EW and Canada data. 58
1.4 Means and standard deviations of period life-expectancies in years 2034 and 2059 for the LC, APC, CBD and M7 mortality models, EW and Canada data. 58

2.1 Z-scores and p-values of Geweke's diagnostic for parameters μ, σ_x^2 and σ_y^2 and for latent states x_6, x_{16} and x_{26} of the Gaussian State Space model. 75
2.2 Posterior summary for parameters μ, σ_x^2 and σ_y^2 and for latent states x_6, x_{16} and x_{26} of the Gaussian State Space model. 77
2.3 Z-scores and p-values from Geweke's diagnostic for parameters of the random walk, $\boldsymbol{\delta}$ and V_ζ, of the B-CBD model, EW data. 85
2.4 Posterior summary of the B-CBD model, EW data. 87
2.5 Cross-correlation matrix of the MCMC series for the parameter sub-vector $\left(\delta_1, \delta_2, \sigma_1^2, \sigma_2^2, \sigma_{12}, \kappa_{2009}^{(1)}, \kappa_{2009}^{(2)}\right)$ of the B-CBD model, EW data. 88
2.6 Predictive distributions of survivor indices for EW males aged 65 in 2010 after $t = 20$ years, between the CBD, PC-CBD and PU-CBD projections. 93

3.1 Rejection rates of the statistical tests at a 5% significance level for the standardised residuals matrices of the LC and CBD models, EW data. 99

3.2 Least squares estimate of the autoregressive matrix, $\mathcal{A}$, from empirical residuals matrices of the LC and CBD models, EW data. 105

4.1 Summary of posterior distributions for the parameters of the BA-LC model and p-values of Geweke's convergence diagnostic, EW data. . . 127
4.2 Summary of posterior distributions for the parameters of the BA-LC model and p-values of Geweke's convergence diagnostic, Canada data. 128

5.1 Summary of posterior distributions of the parameters of the BA-CBD model and p-value of Geweke's convergence diagnostic, EW and Canada data. 151

6.1 Mean rejection rates of statistical tests, at a 5% confidence level, over the posterior distribution of the standardised residuals matrix of the BA-models for EW and Canada data. 162
6.2 Deviance Information Criteria for the Proposed Models 170
6.3 Summary of period life-expectancy distributions at age 65 in calendar years 2034 and 2059 for the BA-CBD and BA-LC models, EW and Canada data. 180

Introduction

Background

Longevity indices have globally increased impressively during the past century in the majority of the developed countries and this trend is expected to continue. Mortality data are usually collected, summarised and released by the local national statistical authorities. The Office for National Statistics (ONS) is the corresponding authorised body in the UK. According to some of the most recent released ONS projections, cohort life expectancy at birth, an index calculated by allowing for mortality improvements over time, is estimated at 90.6 years in 2012, at 94.3 years for 2037 and at 98.0 for 2062, for the UK males population. New-born females are estimated to live 93.9 years if born in 2012, 97.3 years if born in 2037 and 100.7 years if born in 2062 (ONS, 2013b).

The standard approach for the calculation of mortality related indices is based on an artificial statistical tool, known as a *life table*. Life tables are usually constructed separately for different sexes since females experience lighter mortality over their life. Standard textbooks with extensive information about life tables and their usage in practical applications may be found by Pitacco et al. (2009) or Dickson et al. (2013). Commonly they include metrics and indices obtained after *graduating* or *smoothing* raw mortality data. Such metrics are the *central death rate* or the *mortality rate* for age x, m_x and q_x, respectively. These might be used to produce certain measures of the health status of a population, such as the life expectancy for all given age-specific mortality rates. In contrast to cohort objects, period life tables assume that mortality is constant over time and only varies across the ages included in the life table. Although they give a static representation of mortality, they provide a useful tool for summarising the experience of a population over a given period, and, for past years, provide an objective means of comparison of the trends in mortality over time. Calculations of mortality indices depend on the methodology used by the issuing body. For example, in some cases the ONS computes period life expectancies for its constituent countries using three-year rolling averages, based on deaths registered in calendar years and mid-population estimates using the Chiang method (Chiang,

1984, ONS, 2014). The development of period life expectancies at birth and at age 65 for the England and Wales (EW) males and females as estimated by ONS is shown in Table 1 (ONS, 2013a). The table displays that life expectancy is rising for both sexes and at both ages. However, the improvement is faster for life expectancy at age 65 than at birth. Characteristically, the average improvement of life expectancy at birth is 3.6% for EW males over the last 3 decades, but the corresponding figure for life expectancy at age 65 is 12%. Similarly, the improvements are stronger for males than females. For example, the average improvement of life expectancy at 65 is 7.07% for females in contrast to the 12% of EW males. On the other hand, females enjoy systematically much better longevity indices than men.

	At birth		At age 65	
Year band	Males	Females	Males	Females
1980-82	71.04	77.00	13.04	17.02
1990-92	73.36	78.87	14.24	18.01
2000-02	75.86	80.53	16.02	19.12
2010-12	78.96	82.79	18.31	20.88

Table 1: Period life expectancies at birth and at age 65 for EW males and females as estimated by the ONS.

Beyond the national statistical authorities' releases, a global mortality database is maintained by the University of California, Berkeley (USA) and the Max Planck Institute for Demographic Research (Germany). The Human Mortality Database (HMD) includes raw statistics, graduated rates and estimated mortality indices for various age groupings and years for the majority of developed countries. Unlike to National Statistics which are often broken down regionally, the HMD data consist of estimates for the whole population of the specified country. A detailed description of the exact methods applied by the HMD may be found in Wilmoth et al. (2007). Figure 1 depicts the HMD calculated period life expectancies at birth of the EW males and females for the period 1841-2011. Beyond the dominance of females' life expectancy over the whole period, it is evident that the 1918-19 influenza hit males much harder than females. Moreover, males lost a significant amount of their life expectancy during World War II. During the succeeding years the difference in life expectancy of males and females increased before the males' index started to accelerate during the last 30 years, as also shown in Table 1.

The experienced mortality improvements in the developed countries are highly responsible for the present ageing of the populations. The concern regarding the health and social security of these populations has attracted increasing attention (Booth et al., 2006). On the other hand, according to several studies, mortality projections for low mortality countries have tended to underdetermine the realised improvements (Lee and Miller, 2001). Such inaccurate estimates are of high interest in Insurance and Actuarial Science.

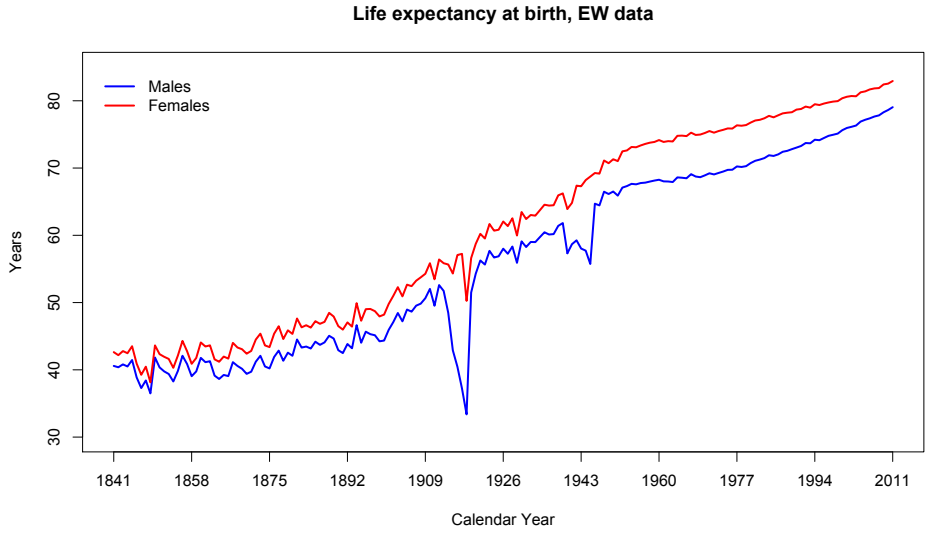

Figure 1: Period life expectancy at birth of EW males and females for the period 1841-2011 as estimated by the Human Mortality Database.

From the point of view of pension and annuity providers, continuously decreasing trends in mortality rates hide the danger of facing capital insufficiency. A typical example is met in the case of Equitable Life Assurance Society (ELAS), which in 2000 failed due to unhedged exposure to both longevity and interest rate risk. ELAS' tactic included providing policies with Guaranteed Annuity Rates, which were priced under certain interest and mortality rate assumptions. When interest rates and life expectancy turned out to be, respectively, lower and higher than anticipated, the business failed to meet its clients demands for lump sum transfers and the firm was closed to new business (Blake et al., 2006). Apart from ELAS' poor management of interest rate risk, the significant exposure to longevity risk led to the acknowledgement that mortality is a key risk factor which cannot be ignored. The life insurer on the other hand is always interested in having an accurate estimate of the force of mortality since his reserves depend and are based on it. Finally, the exposure of institutions and governments to such low mortality rates has initiated financial instruments based on future mortality statistics, which are to be utilised so that the relative risk is hedged (Biffis and Blake, 2009).

The mortality indices inserted in a life table are obtained after smoothing the data under some appropriate *mortality model*. Mortality data are usually consist of observed deaths and population estimates. A mortality model is a mathematical equation which describes the evolution of mortality rates relative to some factors that the process depends on. Historically, mortality modelling falls within the fields of Demography and Actuarial Science. However, it embraces fields such as Sociology, Medicine, Fiscal Finance, as well as general Finance, and therefore is turned into a hugely interdisciplinary subject. Therefore, a globally true model would require several different covariates in order to mimic reality accurately. The statistical approach to the problem is the one which makes use of past trends and focuses in projecting them to the future.

Setting

The HMD data are given in matrix form with entries corresponding to deaths and exposures-to-risk for integer values of age and year. The ages, x, are usually recorded as age last birthday and henceforth aged x or just (x) will have that meaning. A one year period runs from time t to time $t+1$. The latter period can also be called *calendar year* t. Throughout the thesis we denote m and n the number of ages and years of the data, respectively. The exposures have the meaning of an average population estimate for some age-year cell. They could also be read as the total waiting time of the population. Lastly, an individual who is aged x in year t, is assigned to belong to the cohort $c = t - x$, where c indicates the approximate year of birth.

Given the above specifications, it is useful to define:

- $D(x, t)$ as the number of deaths of individuals (x) during year t,
- $E(x, t)$ as the central exposures-to-risk of individuals (x) during year t.

The imminent estimate for the mortality rate is the crude death rate, which is given simply as:

$$\widetilde{m}(x, t) = D(x, t)/E(x, t).$$

Figure 2 shows the observed log-mortality surface and the age-specific crude log-death rates for the EW data. Based on these plots one can make several observations.

- Mortality appears to improve for all ages across the whole observation period.
- The rate of improvement has been considerably different at different ages. For example, it is visually evident that the log-rates of at age 60 have been declining much faster compared the corresponding figures for age 89. If the observed rates were plotted over a greater time period than that of the bottom panel of Figure

2, it would also be possible to see that improvement rates have been varying over time, with decades of significant improvements being followed by decades of stabilisation and vice versa.

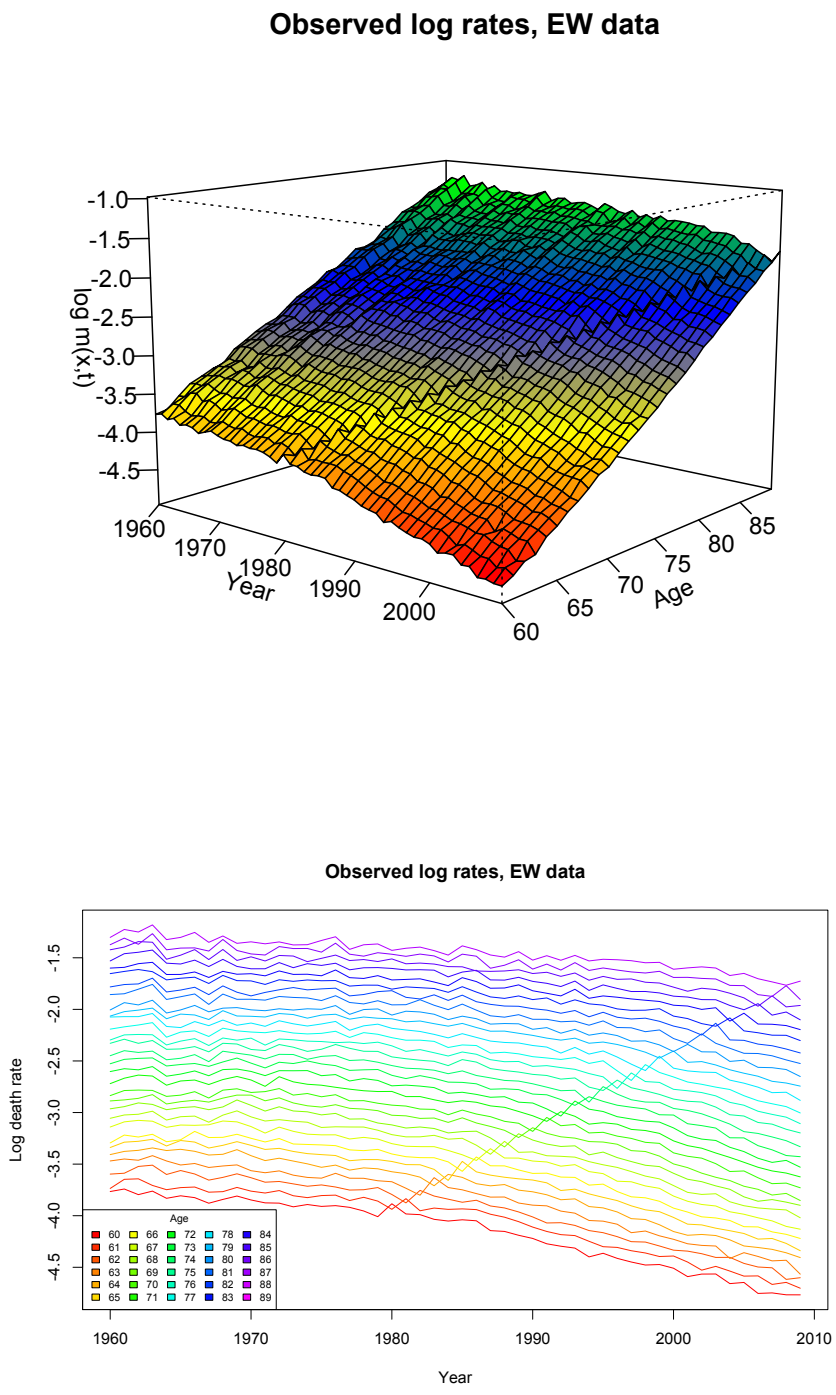

Figure 2: Crude log-mortality surface and age-specific crude log-death rates for the EW data on the top and bottom panels, respectively.

- There are diagonal patterns indicating that the mortality experience is a shared characteristic between individuals with the same year of birth or simply the cohort effect whose significance has been well recognised (Willets, 2004, Richards et al., 2006).

To sum up, mortality rates for EW follow a declining pattern and, generally, decline rates are different for different age groups and during different time periods. Moreover, beyond the intuitive factors of age and time, there is an obvious interdependence relationship indicated by the cohort effect.

Methods

Actuarial models of mortality and life tables have been developed over centuries under deterministic or scenario-based technologies. Applications involving mortality statistics are met as early as the start of the 18^{th} century, and naturally, various methods and techniques have been employed. According to the first laws of mortality and during the early stages of the relevant work, it had been common to project mortality solely based on age. However, under more modern techniques, the effect of calendar year is also well acknowledged.

A fundamental problem is to estimate a quantity known as the *force of mortality* or alternatively, the probability of dying for individuals belonging in a particular group, at some future time instance. Focusing on contemporary methods, in the UK it has been common to use *reduction factors* or *improvement rates* (Renshaw and Haberman, 2000). Given a present estimate of the quantity of interest, the use of reduction factors consists of multiplying that estimate with some appropriate function. The name follows from the fact that future mortality is expected to reduce or improve. Most often the mortality rate is assumed to reduce by a fixed percentage annually and the percentage is determined by past observations. These models do not commonly allow for an explicit quantification of uncertainty, and the parameters are typically based on a mixture of recently observed trends and expert opinion (Wong-Fupuy and Haberman, 2004). Although Li et al. (2010) show that the framework may appropriately get accommodated for practical actuarial tasks, there have been evidence to indicate systematic underestimation of mortality improvements under this approach (CMI, 2007b).

As already mentioned the raw data are graduated in order to be summarised. Smoothing out mortality data is sometimes the primary concern of the mortality model. Firstly, it naturally sensible for mortality rates to be increasing with age for a fixed observation year. More practically, projections of mortality measures are used in pricing products and calculating reserves, any irregularities would then be passed to these

calculations. As observed mortality rates are inherently noisy, due to measurement errors and the estimated quantities involved, graduation arises as a natural need.

A common non-parametric method of graduation is by using splines. A spline is a piecewise defined function of polynomials. Each polynomial is determined for an interval denoted by two knots. B-splines constitute a family of curves which are zero outside a given short interval. They make the estimation of the parameters of the model much simpler and are a usual choice of basis for a spline model. Eilers and Marx (1996) introduced parameter estimation of a model with B-splines basis through penalised likelihood, leading to the development of P-splines. The imposed difference penalty is there to avoid over-fitting and it is chosen subjectively. Currie et al. (2004) used a basis of bi-dimensional P-splines on which the multidimensional mortality data are regressed. The penalty is imposed for each cell and depends on the model parameters. The method can also be extended by weighting the penalties, so that the knots can be increased without significant alternation to the smoothness of the fit (Pitacco et al., 2009). In practice, the P-splines approach assumes some degree of smoothness across all dimensions: age, period and cohort. However, excessive smoothing in the period dimension can lead to systematic over- or under-estimation of mortality rates, as the model fails to capture what may be genuine environmental fluctuations from one year to the next that affect all ages in the same direction (Cairns et al., 2007). Nevertheless, the method proves to provide a good overall fitting and is a highly popular means of graduation (CMI, 2007a). On the other hand, the choice of the penalty relies on the future pattern of mortality and the degree of smoothing in empirical applications depends on the variability of the observed death rates. Moreover, forecasting is a direct extrapolation of the fitted curve and therefore does not enable direct stochastic simulation of future rates.

Nowadays, it is widely acknowledged that improvements in mortality have been accomplished in a non-deterministic fashion. Hence, the acknowledgement of mortality under a stochastic framework has been the most modern approach in academic and industrial modelling. Stochastic mortality models usually express key indices, such as the force of mortality or the mortality rate as a stochastic process which is estimated by the past development and determines the future dynamics. Stochastic mortality models might be divided into two broad categories identifying the two different underlying modelling approaches. The first approach develops stochastic trend models that are extrapolative in nature, by focusing on the time-series properties of mortality, and targets building models that describe the process as well as possible in a statistical sense. The foundations of modelling mortality using time series can be found in McNown and Rogers (1989), whereas one of the most influential models was developed by Lee and Carter (1992). The other approach is more based on the mathematical

aspects of modelling, with less emphasis on statistical model validation, and takes a pricing and risk management perspective analogous to that within financial derivatives literature. The models share common characteristics with interest rate models and typical examples include work developed by Milevsky and Promislow (2001) and Dahl (2004). These models tend to be simpler in nature, but make it possible to assess the potential for dynamic hedging of mortality risk (Cairns et al., 2008). They provide the framework for determining the price of risk and the transformation from the historical or real world probabilities, to the risk neutral, or pricing, probabilities. They usually refer to single age cohorts, and for allowing multiple age modelling the dependence across ages must be considered efficiently.

The modelling of mortality as a stochastic process is a relative recent development, which is supported by the uncertainty related to mortality development and the associated need to better quantify and manage this uncertainty. Lately, various models for the evolution of mortality have been proposed by several authors, and by now there exists a vast literature on the subject. Extrapolative time-series models express age-specific mortality rates as functions of time depending on parameters which are estimated from historical data. A model for forecasting is therefore built so that predictions for the index of interest are made. The sound theoretical framework of time-series also results in relative parsimonious models, whereas forecasts are based on the observed long-term dynamics. Finally, quantification of uncertainty in forecasts is achieved in the form of probabilistic confidence bands.

The approach of Lee and Carter was subsequently followed by several authors and has yielded a vast variety of models. Representatives of the generalised Lee-Carter (LC) family of models could include members given by Brouhns et al. (2002), Renshaw and Haberman (2003b), Czado et al. (2005), Renshaw and Haberman (2006) and Delwarde et al. (2007), amongst others. More from an actuarial perspective, compared to Lee and Carter's demographic approach, Cairns et al. (2006a) suggested a two-factor stochastic mortality model, based on a stochastic version of models introduced by Perks (1932). The resultant Cairns-Blake-Dowd (CBD) model, and several direct extensions of it, are presented by Cairns et al. (2009) where members of both the generalised LC and CBD families are compared based on qualitative and quantitative criteria, after being fitted to the EW and USA males populations. Subsequently, their study continues with formal goodness of fit evaluation and backtesting of the implemented models (Dowd et al., 2010a,b), and concludes by examining forecasts of the dominant considered structures (Cairns et al., 2011a). The above series of work was the basis for developing the LifeMetrics index by the Pensions Advisory Group of JP Morgan Chase (Coughlan et al., 2007). Building on the advantages of both the LC and CBD models, Plat (2009) introduces an extended interface between them.

Plat's model combined the simple multi-factor structure of the CBD, where the age interaction terms are simple functionals of the ages of the data-set, with the isolated age effects of the LC model. The outcome is a multi-factor model which is able to describe mortality across the whole age span. Plat's model is widely used in industry by Insurance and Reinsurance companies. For example, Swiss Re, the world's second biggest Reinsurer, is using a slightly modified version of the model. Their model was developed as part of work developed by Haberman and Renshaw (2011), where Plat's modelling ideas were extended and where a comparative study similar to that of Cairns et al. (2009) was conducted.

Objectives

The present Thesis is based on discrete-time stochastic mortality models that have been the mainstream trend in research during the past 20 years. Throughout that period several underlying frameworks have been proposed and researchers have experimented with various types of models. During that era of development several requirements for good stochastic mortality models have been set. Some of them are summarised below:

- Mortality rates should be positive.
- The model should be consistent with historical data.
- Long-term dynamics under the model should be biologically reasonable.
- Forecast levels of uncertainty and central trajectories should be plausible and consistent with historical trends and variability in mortality data.
- The model should be relatively parsimonious.
- It should be possible to use the model to generate sample paths and calculate prediction intervals.
- The structure of the model should make it possible to incorporate parameter uncertainty in simulations.
- Parameter estimates should be robust relative to the period of data and range of ages employed.
- Model forecasts should be robust relative to the period of data and range of ages employed.

The procedure of developing improved models has led researchers to creating models of increasing complexity with no evidence of significant improvement in the out-of-sample forecasts. This Thesis suggests an alternative framework where the local mortality effect is appreciated, by employing a parsimonious multivariate process for modelling the latent residual effects of a simple stochastic mortality model as dependent rather than conditionally independent variables. The extension introduces serial dependence within the cells of the data-set by relating the residual terms through a simple vector autoregressive model. The method is applied to the structures provided by the two most popular stochastic mortality models in the literature, namely the LC and CBD models. The work is based on the EW and Canada males populations during the retirement ages of 60-89 years and for years 1960-2009 taken from the HMD. The empirical residual estimates of the original LC and CBD models allow for identifying the interdependencies within the given populations, and lead to the determination of the appropriate residuals model. However, the differences between the two models as benchmarks are also stressed, as the two-factor structure of the CBD model is beneficial in terms of explaining the variability within mortality data. Under the Bayesian paradigm, samples from the joint posterior distribution of the latent states and parameters are obtained by developing Markov chain Monte Carlo (MCMC) algorithms. Along with the resultant short-term dynamics, we also examine the impact of the joint estimation in the original long-term dynamics of the benchmark models. The Bayesian solution aids in recognising the different levels of uncertainty for the two naturally distinct type of dynamics across populations. The forecasted rates, mortality improvements, and more mortality dependent metrics under the developed models are compared to those produced by their benchmarks and other standard stochastic mortality models in the literature.

Layout

The Thesis is organised as follows. Chapter 1 presents the construction of the main indices for measuring mortality and the dominant assumptions about them within the literature. Additionally, the LC and CBD models, which serve as benchmarks for the development of the suggested modelling framework, along with related extensions are illustrated, estimated and projected. Chapter 2 discusses the principles underpinning Bayesian modelling and the intuition behind the associated MCMC estimation algorithms. Hybrid algorithms are developed for estimating a Gaussian State Space model and a Bayesian version of the CBD stochastic mortality model. Chapter 3 justifies the intuition for developing the joint residuals modelling within the mortality context by investigating the properties of the estimated state variables for the EW population. Subsequently, the suggested residuals model is presented and we conduct

an empirical analysis based on the standardised mortality residuals of the LC and CBD models. Equipped with the results drawn from Chapter 3, Chapters 4 and 5 apply the proposed modelling framework within the structures of the LC and CBD models, respectively. Therein, the suggested models are presented in some detail and the MCMC estimation schemes are fully developed. Finally, Chapter 6 examines the resultant dynamics, the goodness-of-fit, the forecasting properties and the robustness of the suggested models.

Chapter 1

Mortality Modelling

In this Chapter the main concepts underlying mortality modelling are introduced. We define the principal indices of interest and state the main modelling assumptions. Early models for mortality rates are presented and the various stochastic contemporary approaches are mentioned. We focus on two families of stochastic mortality models which are extended in the following Chapters of the Thesis. Firstly, the Lee-Carter family is considered by detailing the original Lee-Carter (LC) model (Lee and Carter, 1992) and its cohort enhanced version, Age-Period-Cohort (APC) model (Currie et al., 2004, Cairns et al., 2009). Secondly, we demonstrate the Cairns-Blake-Dowd (CBD) family through the original CBD model (Cairns et al., 2006a) and one of its cohort forms, model M7 (Cairns et al., 2009). The models are implemented under common assumptions, by incorporating the same idiosyncratic source of risk, and the models are thus compared in a coherent manner. Finally, the chosen mortality models are compared in terms of goodness-of-fit and compliance with the underlying modelling assumptions. Their direct forecasted rates are compared according to their family classification, and relevant mortality derivative quantities are examined across all presented models.

1.1 Introduction

This section formally sets the foundations for the construction of the most important indices for modelling mortality. As initial examples, the so-called early models, also known as *mortality laws*, are developed. Furthermore, the most widespread assumptions about mortality modelling accepted in the academic literature are stated. Lastly, key actuarial and demographic metrics, useful as means of model comparison, deriving from mortality rates are defined.

1.1.1 Mortality Rates

Mortality modelling fundamentally involves examining the behaviour of future lifetime distributions. Probabilistic quantities based on such distributions allow for a comprehensive description of the process, and serve as the building blocks of the modelling framework. Under the provided framework, the future lifetime distributions are analysed in terms of integer values of age, x, and calendar year or time, t, under the convention that the age is accounted as age last birthday.

Define $\mathcal{T}(x,t)$, a continuous random variable, indicating the future lifetime of a life aged x at time t. The life will die at age $x + \mathcal{T}(x,t)$ in year $t + \mathcal{T}(x,t)$. Imposing an underlying distributional assumption for $\mathcal{T}(x,t)$ yields a survival model. The distribution function, and the probabilistic complement, of $\mathcal{T}(x,t)$ give rise to basic probabilities of interest.

Definition 1.1.1. (*Survival Function*)
The survival function, $\mathcal{S}(x,t) = \mathbb{P}\left[\mathcal{T}(x,0) \geq t\right]$, represents the probability that the life aged x at time 0 will survive to time t.

The probabilistic complement of the survival function, that is, the distribution of $\mathcal{T}(x,t)$, yields the probability that the life aged x at time 0 will die before age $x+t$.

$$F(x,t) = 1 - \mathcal{S}(x,t) = \mathbb{P}\left[\mathcal{T}(x,0) < t\right].$$

Moreover, it is sensible to assume that:

$$\lim_{t\to\infty} \mathcal{S}(x,t) = 1 - \lim_{t\to\infty} F(x,t) = 0,$$

for fixed age last birthday x, meaning that death eventually is certain. In practice, an upper age bound is applied, usually a number ω between 120-130, indicating a limiting age.

The probability of a life dying or surviving at some future time instance is directly

derived by the distributions just defined. The one-year ahead such probabilities are of particular interest.

Definition 1.1.2. (*Annual Mortality Rates*)
The annual mortality rates, $q(x,t)$ and $p(x,t)$, are defined as:

$$\begin{aligned} q(x,t) &= \mathbb{P}\left[\mathcal{T}(x,t) \leq 1\right] \\ p(x,t) &= \mathbb{P}\left[\mathcal{T}(x,t) > 1\right], \end{aligned}$$

represent the probabilities that the life aged x at time t will die and survive throughout the next calendar year, respectively.

The above quantities are defined for all real ages x and years t. However, they are usually quoted for integer x and t. They are obviously complementary and under a stochastic framework they are only observable after time $t+1$. Their annual form allows for a concrete interpretation of the forthcoming annual risk of dying.

1.1.2 Force of Mortality

In order to monitor the future lifetime of (x), a measure of the instantaneous risk of death is required. In an actuarial context this is known as the *force of mortality*. For its construction consider (x) with associated $\mathcal{T}(x,t)$. The probability that this life will die in the infinitesimal interval $(x, x+\delta)$ is given by the conditional probability:

$$\mathbb{P}\left[x < \mathcal{T}(0,t-x) \leq x+\delta | \mathcal{T}(0,t-x) > x\right].$$

The average of this probability over the length of the interval, δ as δ tends to 0, yields the definition of the instantaneous death rate or force of mortality of (x) on year t.

Definition 1.1.3. (*Force of Mortality*)
Given (x), with future lifetime distribution $\mathcal{T}(x,t)$, the force of mortality of (x) at time t is defined to be the quantity:

$$\mu(x,t) = \lim_{\delta \to 0+} \frac{\mathbb{P}\left[x < \mathcal{T}(0,t-x) \leq x+\delta | \mathcal{T}(0,t-x) > x\right]}{\delta}.$$

If the force of mortality is known, the probability of (x) at time t dying in the interval with length δ is approximately $\delta \times \mu(x,t)$. It is a fundamental quantity since it is intimately related to the survival function of the future lifetime random variable.

Note that:

$$\begin{aligned}
\mu(x+s,s) &= \lim_{\delta\to 0+}\frac{F(x,s+\delta)-F(x,s)}{(1-F(x,s))\times\delta}\\
&= \frac{1}{1-F(x,s)}\frac{\partial F(x,s)}{\partial s}\\
&= -\frac{1}{\mathcal{S}(x,s)}\frac{\partial}{\partial s}\mathcal{S}(x,s)\\
&= -\frac{\partial}{\partial s}\log\left(\mathcal{S}(x,s)\right).
\end{aligned}$$

Integrating the above relationship yields:

$$\mathcal{S}(x,t)=\exp\left\{-\int_0^t \mu(x+s,s)\mathrm{d}s\right\}, \tag{1.1.1}$$

under the boundary condition $\mathcal{S}(x,0)=1$, for any given x.

Therefore, $\mathcal{S}(x,t)$, as given in Definition 1.1.1, is directly connected to the force of mortality. $\mathcal{S}(x,t)$ is also known as the *survivor index* and represents the proportion of a large population of (x) 's on time 0 who survive to age $x+t$ on time t (Cairns et al., 2008). Consequently, the relevant relationship between the force of mortality and the mortality rate, $q(x,t)$ might be derived.
That is:

$$q(x,t)=1-\exp\left\{-\int_t^{t+1} \mu(x-t+s,s)\mathrm{d}s\right\}, \tag{1.1.2}$$

and indicates the probability of (x) on time t dying before time $t+1$.

The behaviour of the force of mortality may also be summarised by the central death rate over the age $(x,x+1)$, which is usually denoted by $m(x,t)$. The central death rate is defined as the age-continuous weighted arithmetic mean of the force of mortality over age x to $x+1$ (Pitacco et al., 2009). Although the central death rate has its roots in Demography since it is crucial for the relevant population projections, it has also been a standard measure of mortality within Actuarial Science too.

Considering the above discussion, it follows that the force of mortality fully describes the future lifetime distribution. Therefore, it is critical to model it efficiently and indeed, a vast variety of models have been proposed for it. According to the early mortality laws, presented in the next section, the force of mortality was assumed to be given from some deterministic parameterisation of age on the log scale, whereas time is just ignored. This approach results in static mortality models, such that $\mu(x,t)\equiv\mu(x)$, whereas a deterministic model implies that the force of mortality is

given as some specified function of both x and t (Cairns et al., 2006b). In contrast, the contemporary approach, and the one that the current Thesis is based on, is that subsequent realisations of the force of mortality are a series of random variables, so that it is of a stochastic nature itself. However, the mortality laws give a cunning insight into the nature of the process' general pattern. Moreover, the majority of the most popular stochastic mortality models are just sophisticated extensions of those early laws.

1.1.3 Mortality Laws

Several parametric functions aiming to describe the mortality evolution have been developed over time. Those presented here assume only age as covariate, but incorporating time as an argument of the respective function is also possible. The simplest of these laws are only suitable for specific age groups by exploiting, for example, the observation that mortality rates are linear on the log-scale for older ages. Models describing mortality across the full age span include additional terms responsible for capturing specified characteristics of the pattern. The initial motivation for them was to describe the cross-section of death probabilities, with primary motivation to smooth data, reduce errors, and create life tables.

The *Gompertz law*, suggested in 1825, is the earliest known force of mortality model, stated as:

$$\mu_x = \alpha e^{\beta x}.$$

The *Makeham law*, proposed in 1867, is another fundamental model, given by:

$$\mu_x = \gamma + \alpha e^{\beta x},$$

for α, β, γ non-negative parameters in both cases.

The Gompertz law essentially states that the logarithm of the force of mortality is linear with age, along with some offset contribution. In the case of the Makeham law, we can interpret the term $\alpha e^{\beta x}$ as the age dependent mortality, whereas the γ term represents the risk of death which is independent of age.

It is intuitively reasonable that mortality is driven by different factors at different ages. This was the motivation for using multi-factor models, where each factor captures the behaviour of the pattern at different parts of the curve. Thiele proposed in 1872 a model including three factors. The first factor represents the mortality at infancy and childhood, the second factor is responsible for capturing the mortality behaviour during adulthood, and the last factor describes the senescent mortality, that is, mortality of the elderly. The sum of the components describes the mortality

pattern across the entire age span, according to the following equation:

$$\mu_x = \alpha_1 e^{-\beta_1 x} + \alpha_2 e^{-\beta_2 (x-\eta)^2} + \alpha_3 e^{\beta_3 x},$$

where all parameters are non-negative.

It is observed that all three models presented so far might be considered as members of a general family of models. Forfar et al. (1988) extensively describe a variety of graduation techniques, and define the class of generalised Gompertz–Makeham(GM) models, given as:

$$GM(r,s) = \mu_x = \sum_{i=1}^{r-1} \alpha_i x^i + \exp\left[\sum_{j=0}^{s-1} \beta_j x^j\right].$$

Under the convention that for $r = 0$ or $s = 0$ the corresponding polynomial or exponential term is ignored, it can be seen that the Gompertz law is given as $GM(0,2)$, whereas the Makeham law arises for $GM(1,2)$.

Under a different point of view, Perks (1932) suggested the logistic family of curves, expressed as:

$$\mu_x = \frac{\alpha_1 e^{\beta x} + \gamma}{\alpha_2 e^{-\beta x} + \alpha_3 e^{\beta x} + 1}$$

which includes the Makeham model for $\alpha_2 = \alpha_3 = 0$, a case which is considered as a member of the GM family. Various such mortality models have been previously proposed and they are not exhaustively focused in the force of mortality. For example, Beard (1971) suggests modelling the mortality rate q_x, whereas Heligman and Pollard (1980) model mortality over the whole life span with a class of formulae, of which the first targets the odds of death $\phi_x = q_x/p_x$, and states that:

$$\phi_x = A^{(x+B)^C} + D \exp\left[-E\left(\log x - \log F\right)^2\right] + GH^x,$$

where A, B, C, D, E, F, G and H are non-negative constants (Pitacco et al., 2009). Early models for graduation are summarised in Forfar et al. (1988), and examples of such models under a more contemporary approach are given by Pitacco et al. (2009). Although the above models are no longer used under their presented forms, they provide solid foundations for developing more complicated structures under different frameworks. In the simplest case, their parameters might become functions of time, leading to a deterministic model which can be extrapolated to obtain future mortality estimates.

The well recognised stochastic nature of mortality however, would suggest that all, or some of those parameters, are themselves stochastic processes, so that their dynamics

need to be specified. For example, Dellaportas et al. (2001) introduce a Bayesian variant of the Heligman-Pollard model, whereas the model suggested by Cairns et al. (2006a) is a stochastic version of the Perks logistic model. Lastly, Milevsky and Promislow (2001) assume that mortality has a Gompertz form, for which the offset is modelled using a mean-reverting continuous-time diffusion. All the above examples are direct extensions of early laws, something that highlights their importance. The fundamental ideas underlying stochastic mortality, together with the accompanied dominant assumptions in the literature, are introduced in the next section.

1.1.4 Modelling Assumptions

Mortality rates have improved dramatically over the last half of the 20^{th} century (Macdonald et al., 1998). Additionally, the evolution of the process has been very similar across the majority of developed countries (Tuljapurkar et al., 2000). However, it has been noted that those improvements have been accomplished in a non-deterministic fashion (Currie et al., 2004). In particular, the rates of improvement vary notably over time and moreover, the improvements have varied significantly between different age groups. As a result, future mortality improvements cannot be forecasted with any degree of precision (Cairns et al., 2006b).

In that sense, the models introduced in the previous section are useful but inadequate due to their deterministic framework. From a demographic and actuarial perspective, it was traditional to assume some deterministic function f_μ, indicating the force of mortality such as the early laws, according to which:

$$\mathbb{P}\left[t < \mathcal{T}(x,0) < t+\delta\right] \approx \delta \times f_\mu(x+t), \tag{1.1.3}$$

where $\mathcal{T}(x,0)$ is the random residual lifetime of a life aged x at time 0. Under the above specification, individuals of the same age share the same force of mortality regardless of the time instance. Time might be incorporated in f_μ so that it acquires a dynamic nature.

In contrast, the stochastic nature of mortality focuses in assuming that the modelled indices are themselves stochastic processes. The latter is done by assigning random dynamics to them or the parameters that construct them and the stochastic intensities are thus interpreted as instantaneous death probabilities conditional on specified information (Biffis, 2005).

To distinguish cases, consider the random variable $\mathcal{T}(x,0)$ now driven by the random intensity μ_x, and denote $\mathcal{F}_t$ the information generated by μ_x up to and including time

t. The basic concept underlying stochastic mortality is summarised in:

$$\mathbb{P}\big[\mathcal{T}(x,0) \leq t+\delta | \mathcal{F}_t\big](\omega) \approx \delta \times \mu_x(t,\omega), \tag{1.1.4}$$

for some state of the world ω, such that $\mathcal{T}(x,0)(\omega) > t$. Equation (1.1.4) is viewed as the stochastic analogue of (1.1.3), and from an actuarial perspective, emphasises the stochastic variation of the intensity over time, as new information about mortality becomes available (Biffis, 2005).

Such an assertion turns expressions (1.1.1) and (1.1.2) from probabilities to random variables, observable only after their realisation, and this is still an over-simplifying assumption since in reality merely an estimate of them is obtained from a finite amount of data, considerably later than realisation. The required probabilities at time 0 are given by taking expectations under the realised mortality term structure. The argument similarly extends so that probabilities at any time t are derived (Cairns et al., 2006b).

As mentioned in the Introduction, the stochastic mortality models developed so far can be broadly distinguished as continuous and discrete-time. Beyond the examples stated by Milevsky and Promislow (2001) and Dahl (2004), which generally take the form of the early laws driven by diffusion processes, more complex approaches have been found by Biffis (2005) who imposes an affine structure on Dahl's approach. On the other hand, models of the discrete-time approached are summarised in studies by Booth and Tickle (2008), Cairns et al. (2009), Haberman and Renshaw (2011), which are generally based on ARIMA time series methods.

The data distribution supports the use of discrete time models, and usually the modelled index is either $q(x,t)$ or $m(x,t)$. One main assumption, which is usually adopted in the literature under this framework, is that within each and every (x,t)-cell in the data we admit that the force of mortality is constant, so that for $0 \leq i,j < 1$:

$$\mu(x+i,t+j) = \bar{\mu}(x,t). \tag{1.1.5}$$

It is quite intuitive that the instantaneous death rate, $\mu(x,t)$, is close to the central death rate $m(x,t)$ in the middle of the interval, so that $\mu(x+1/2,t+1/2)$ equals $m(x,t)$. That can also be justified more formally, as for example in Forfar et al. (1988). Under assumption (1.1.5) this yields:

$$m(x,t) = \bar{\mu}(x,t). \tag{1.1.6}$$

Now from equations (1.1.2) and (1.1.6) it follows that:

$$q(x,t) = 1 - \exp\left[-\bar{\mu}(x,t)\right] = 1 - \exp\left[-m(x,t)\right]. \tag{1.1.7}$$

The above approximation is felt to be accurate enough and is useful in comparing models that are based on different mortality indices (Cairns et al., 2009).

Under the assumption that deaths occur independently in each age-year cell, given the underlying mortality index, two models are common in the literature. Firstly, the Poisson assumption detailed by Brouhns et al. (2002) states that:

$$D(x,t)|m(x,t) \sim Poi\big(E(x,t)m(x,t)\big). \tag{1.1.8}$$

The Poisson model is a standard actuarial model which along with assumption (1.1.6), leads to a compact likelihood that may be maximised for any particular assigned parameterisation of $m(x,t)$. For a model which focuses on $q(x,t)$, (1.1.8) can still be used, as under assumption (1.1.5), the relationship (1.1.7) might be used.

Another popular actuarial model about the occurrence of deaths arises under the Binomial distribution. Index $q(x,t)$ might be viewed as the parameter of a Binomial distribution, so that:

$$D(x,t)|q(x,t) \sim Bin\big(E^i(x,t), q(x,t)\big). \tag{1.1.9}$$

In this case the exposures take the form of the population at risk at the start of the calendar year t. Given data for the central exposures and deaths, the initial exposures are retrieved by:

$$E^i(x,t) = E(x,t) + D(x,t)/2.$$

The above approximation implies that deaths occur on average half-way through calendar year t for all ages x. Both of the above stated assumptions are popular and natural sources of randomness within mortality modelling. Practical applications show that there are very small differences between the results of the approaches.

1.1.5 Mortality Metrics

The usual way to measure and compare mortality is through indices such as the graduated central death rates, $m(x,t)$, or the corresponding mortality rates, $q(x,t)$. The outcomes of a mortality model are usually summarised in life-tables. Life tables are published by the appropriate authorities in the majority of developed countries. For example, the UK Government Actuary prepares periodically the English Life Tables published by the ONS. In the USA, the Division of Vital Statistics of the National

Center for Health Statistics prepares the US Life Tables (Dickson et al., 2013).

Period life-tables tabulate graduated mortality statistics for integer ages, fix an initial radix for the population under consideration and illustrate the mortality experience over a certain period of time. Assuming that the members of the population will experience the particular age-specific mortality rates for that period throughout their lives and that the population size is stationary, there have been several actuarial methods developed for calculating quantities of interest such as life expectancies and survival probabilities. For example, as already mentioned, the ONS uses the Chiang method in the construction of their period calculations. The Chiang method for life-table calculations is implemented in the Demography package in R (Hyndman et al., 2014). Although period life table calculations make no allowance for any mortality improvements, they provide solid means of comparison between the outcomes of different mortality models.

A standard measure of health of the population in study is the life-expectancy at age x. Under the static framework of the period life-table, the future lifetime random variable, defined in Section 1.1.1, takes a time invariant form such that $\mathcal{T}(x,t) \equiv \mathcal{T}_x$. It is then a standard result that the density of $\mathcal{T}_x$ is given as:

$$f_x(t) = \mathcal{S}_x(t)\mu_{x+t},$$

where $\mathcal{S}_x(t)$ is the survival function of $\mathcal{T}_x$ and μ_{x+t} is the underlying force of mortality. The life-expectancy of (x) is then taken as the expectation of $\mathcal{T}_x$:

$$L_x = \int_0^{\omega-x} t f_x(t)dt = \int_0^{\omega-x} \mathcal{S}_x(t)dt = \sum_{j=0}^{\omega-x} \int_j^{j+1} \mathcal{S}_x(t)dt \approx \frac{1}{2} + \sum_{j=1}^{\omega-x} \mathcal{S}_x(j),$$

where the last step follows if the integrals of the previous summation are approximated with the trapezoidal rule, and ω is some presumed limiting age. Given the information of the period life-table it is simple to calculate the expression given above, since the extraction of the relevant survival function is taken by a straightforward multiplication. Each of $\mathcal{S}_x(j)$ represents the probability for the life (x) to survive for the j next years. Additionally, it is simple to show that the second term of the final equality is also the curtate life expectancy, *i.e.* the integer value of years the life is expected to live. The relationship, beyond being a simple formula for calculating period life expectancies, also yields a basic approximate connection between complete and curtate measures of health.

In practice mortality indices are expected to vary over time. So, period life expectancy does not give the number of years someone could actually expect to live. Cohort life expectancies are calculated using age-specific mortality rates which allow for changes

in mortality in later years, by incorporating an assumption on how future mortality will develop in these future years. Hence, such indices are generally regarded as more appropriate means of measuring the actual number of years a person is expected survive. Given age-specific mortality rates across some period of time and under the modelling assumptions of Section 1.1.4 the survivor index, defined in equation (1.1.1) may also be calculated by a straightforward multiplication by following the appropriate cohort across the diagonal of the two-dimensional life table.

Usually the range of ages of the modelled data-sets used do not include observations up to that limiting age. For example, the data used in the Thesis cover ages up to 89 years old. In these cases, the further rates up to age ω cannot be produced by the simulation of the mortality model and instead, the projected rates are regressed on age and missing rates are obtained by extrapolation. Whenever there is such a requirement for the calculations of the present work, we apply the regression:

$$u(x_0+j,t_0+j)=\text{logit}\big(q(x_0+j,t_0+j)\big)=a+b(x_0+j)+c\sqrt{x_0+j},$$

where x_0 and x_m are the first and last ages included in the dataset respectively, t_0 is the first projection year and $j=0,\ldots,x_m-x_0$, and after estimating the parameters a,b and c, the extrapolated rates are produced by the relationship:

$$u(x_m+j^*,t_0+x_m-x_0+j^*)=a+b(x_m+j^*)+c\sqrt{x_m+j^*},$$

where $j^*=1,\ldots,\omega-x_m$. The method is a modification of a standard demographic approach introduced by Coale and Kisker (1990) to top-up required rates by age. Although the original method suggested by Coale and Kisker (1990) uses a quadratic in age term as the regressor of c, the linear models that resulted from our data-sets seem to be systematically develop a higher R^2 measure with a square root term for c. Hence, we modify the extrapolation formula as shown above.

The survivor index might also be used for mortality depended quantities, useful for measuring the associated degree of model risk. For example, the present value of term annuities payable annually in arrears over a certain number of years is also used to examine quantitatively the outcomes of the mortality models in the following sections. Under a constant interest rate, i, assumption through time, with corresponding discount factor $u=1/(1+i)$, the required annuity price payable to (x) in arrears for a maximum of y years is given as:

$$P=\sum_{t=1}^{y}u^t\mathcal{S}(x,t).$$

The stochastic approach employed gives rise to the future distributions of these quantities which can then be appropriately summarised to compare the outcomes of the different mortality models.

1.2 The Lee-Carter Model

The majority of practical modelling work has been conducted under the discrete time framework as commented in Section 1.1.4. Under the stated, assumptions simple ARIMA models are devised to project the stochastic factors of the mortality models and construct future rates. The model introduced by Lee and Carter (1992) is considered the one that engaged the interface between the two fields. However, ARIMA models had been used previously by McNown and Rogers (1989) in a parameterised mortality context. The original Lee-Carter model was suggested as a demographic alternative to direct life expectancy forecasting. Its dynamics are assumed to be driven by a single random walk with drift model. The simple structure of the model achieves a competitive reflection of the mortality pattern across the whole age span over time.

According to the general framework Lee and Carter suggest, mortality modelling is a two stage procedure. The first step consists of estimating the parameters of the model from the data. Given these estimates, the intermediate step is to estimate the parameters of the underlying stochastic model. Sometimes at this stage, a choice regarding the optimal fitting period is made. The second step consists of projecting the stochastic factor and constructing best mean estimates for future rates, along with the associated confidence bands. The forecasted rates can then be used to calculate several demographic and actuarial quantities, such as those presented in Section 1.1.5.

Since the Lee-Carter (LC) model is a prototype there are several drawbacks associated. However, the intuition behind the model has served as the basis for various further developments and extensions under a stochastic mortality context. Alterations of the model vary from differentiating the underlying statistical assumptions, for example as in Brouhns et al. (2002), to specifying the optimal period for estimating the parameters of the random walk and including additional interaction factors that mortality depends on (Booth et al., 2002).

The present section illustrates the LC model and extensions based on it. The model is examined under different existing frameworks and is fitted to the data-sets of the Thesis.

1.2.1 Description

The LC model triggered a whole methodology in forecasting mortality. Originally, the model served for breaking-down, analysing and forecasting the mortality experience of the US population, across the full age span, regarding years since the start of the 20^{th} century until 1990. Since then it has been turned into the dominant method, at least for the US social security system and its related forecasts (Hollman et al., 2000). Lee and Carter applied a principal components approach to mortality forecasting, which, in contrast to parametric approaches that specify the functional form of the age pattern of mortality in advance, estimates the age pattern from the data. Forecasting is done by the extrapolation of the time related parameters using time series methods.

The model attempts to describe the unobservable central death rate $m(x,t)$ according to the following equation:

$$\log\big(m(x,t)\big) = \beta_x^{(1)} + \beta_x^{(2)}\kappa_t^{(2)}. \tag{1.2.1}$$

In the above formulation, the age parameters $\beta_x^{(1)}$ describe the underlying log-mortality pattern across all ages which is independent of time. The second age related set of parameters, $\beta_x^{(2)}$, indicate the sensitivity of the logarithm of the central rates at age-x to variations in the period related effects $\kappa_t^{(2)}$. Parameters $\kappa_t^{(2)}$ illustrate the change in the level of mortality over time.

The model specification uses the tendency of age-specific death rates to move up and down together over time. This holds because these changes are modelled as if they are driven by a scalar factor $\kappa_t^{(2)}$. Subsequently, this implies that the modelled rates and the resulting errors are perfectly correlated across ages (Cairns et al., 2008). In practice though, the relative speed of decline at different ages varies. However, the model is characterised by its simplicity and transparency. On the other hand, it only includes one factor to model the dynamics of the process and this is often insufficient.

The right hand side of equation (1.2.1) consists only of latent variables and hence the model suffers from identifiability problems. This means that the model produces identical fitting, for example for the following transformations of the parameters:

$$\left\{\tilde{\beta}_x^{(1)}, \tilde{\beta}_x^{(2)}, \tilde{\kappa}_t^{(2)}\right\} = \left\{\beta_x^{(1)}, \beta_x^{(2)}/c, c\kappa_t^{(2)} | c \neq 0, \text{constant}\right\},$$

so that, the $\beta_x^{(2)}$'s and the $\kappa_t^{(2)}$'s are only computable up to some multiplicative constant. The same would have happened if we were to use the transformation:

$$\left\{\tilde{\beta}_x^{(1)}, \tilde{\beta}_x^{(2)}, \tilde{\kappa}_t^{(2)}\right\} = \left\{\beta_x^{(1)} - c\beta_x^{(2)}, \beta_x^{(2)}, \kappa_t^{(2)} + c | c \neq 0, \text{constant}\right\},$$

in which case, the $\kappa_t^{(2)}$'s are calculated up to linear transformation and the $\beta_x^{(1)}$'s are determined up to some linear adjustment. In order to overcome the illustrated identifiability problem, the parameters need to be constrained.

The dominant such constraints across the literature are (Pitacco et al., 2009):

$$\sum_t \kappa_t^{(2)} = 0, \sum_x \beta_x^{(2)} = 1. \tag{1.2.2}$$

Now if we sum the model equation over all years t in hand, we get that:

$$\sum_t \log\big(m(x,t)\big) = n\beta_x^{(1)} \Leftrightarrow \beta_x^{(1)} = \frac{1}{n}\sum_t \log(m(x,t)).$$

Therefore, the first of the constraints yields that for each age x the estimate of $\beta_x^{(1)}$ equals the mean of $\log\left(m(x,t)\right)$ over the years t of the dataset. That point also gives a good initial value for the age effects $\beta_x^{(1)}$ within an estimation algorithm. The second constraint is just an arbitrary sum-to-one constraint, and just like with the first one, there are various suggestions in the literature about it (Cairns et al., 2009).

1.2.2 Estimation

The model might be estimated under various different frameworks and several alternatives have been proposed. Originally, the least squares problem was effectively solved using Singular Value Decomposition (SVD). According to the method presented by Lee and Carter (1992), the $\beta_x^{(1)}$ parameters are set equal to $\hat{\beta}_x^{(1)} = \frac{1}{n}\sum_t \log(\widetilde{m}(x,t))$, where $\widetilde{m}$ denotes the crude estimate as given in the Introduction, and SVD is applied to the matrix of $\left[\log(\widetilde{m}(x,t)) - \hat{\beta}_x^{(1)}\right]$. The first left and right singular vectors give tentative estimates of $\beta_x^{(2)}$ and $\kappa_t^{(2)}$ respectively. To identify a unique set of parameters, the constraints (1.2.2) are then applied to the $\beta_x^{(2)}$ and $\kappa_t^{(2)}$ sets. This leads to an objective function equivalent to maximum likelihood estimation, under the assumption that the errors are homoscedastic and normally distributed. Since SVD is a purely mathematical approximation, the fitted and observed number of deaths may not coincide. Therefore, the original method requires readjustment of the $\kappa_t^{(2)}$'s to fit the actual death rates, treating these derived from the SVD $\beta_x^{(1)}$'s and $\beta_x^{(2)}$'s as known, and the final observation year (also known as jump-off year) rates are taken to be the fitted ones. Lee and Miller (2001) in contrast, adjust the $\kappa_t^{(2)}$'s to the life expectancy at birth and take as jump-off rates the actually observed.

In principle, the assumption of homoscedastic errors is typically violated. This seems to be a shared characteristic regarding the majority of available mortality data (Cairns et al., 2009). An alternative method to the ordinary least squares approach is the

weighted least squares, with the weights being given by the number of observed number of deaths for each cell (Wilmoth, 1993). This has the effect of bringing the parameter estimates close to those derived under the Poisson assumption for deaths. The conditionally independent Poisson model, applied by Brouhns et al. (2002) within the LC framework, has also been a popular choice as mentioned in Section 1.1.4. Towards facing the same issue, Renshaw and Haberman (2003a) work under the Generalised Linear Models framework and assign the deaths an overdispersed Poisson distribution, by including an additional parameter which multiplies the variance, to control the overdispersion of the distribution. Under their approach, the excessive variability is assigned equal for all ages, although this contradicts empirical evidence. Also, the variance controlling parameter is not included in the estimation scheme and therefore the parameter estimates remain unchanged, which also appears inconsistent. Koissi and Shapiro (2006) implement a fuzzy formulation of the LC model to address violations of the constant error variance across age assumption. Li et al. (2009) extend the model by allowing for cell specified heterogeneity by the inclusion of an additional age-specific random variable. Under appropriate assignment for that age-specific variable, the deaths according to their model possess a Negative Binomial distribution with naturally greater variance than mean.

Under the Poisson model for deaths the full estimation procedure is now developed. The log-rates under the LC model are parameterised in terms of age and period effects. As will be illustrated in subsequent sections there is the possibility of including further interaction terms in the equation so that a more complicated model is built. Then the following basic algorithm, allowing for appropriate adjustments, might be used to estimate models in a unified manner. For that, assume that the log-rates of the LC model are written as:

$$\phi(\boldsymbol{\theta}) = \log\big(m(x,t;\boldsymbol{\theta})\big) = \beta_x^{(1)} + \beta_x^{(2)}\kappa_t^{(2)},$$

where $\boldsymbol{\theta}$ is the global parameter vector of the model, which in this case consists of the age and period effects, $\beta_x^{(1)}, \beta_x^{(2)}$ and $\kappa_t^{(2)}$, respectively. Then, the Poisson likelihood is written:

$$\mathcal{L}\big(\boldsymbol{\theta}\big) = \prod_{x,t} \frac{\big[E(x,t)m(x,t;\boldsymbol{\theta})\big]^{D(x,t)}}{D(x,t)!} \exp\big(-E(x,t)m(x,t;\boldsymbol{\theta})\big).$$

The log-likelihood takes the form:

$$\ell(\boldsymbol{\theta}) = \sum_{x,t} \Big[D(x,t)\phi(\boldsymbol{\theta}) - E(x,t)\exp\big(\phi(\boldsymbol{\theta})\big)\Big] + C, \tag{1.2.3}$$

where C is some constant independent of the the parameter vector $\boldsymbol{\theta}$.
The estimation of the components of the parameter vector $\boldsymbol{\theta}$ reduces to maximising the above objective function over the whole parameter space.
Differentiating the function for fixed component θ of $\boldsymbol{\theta}$ yields:

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta} = \sum_{x,t} \left[D(x,t) - E(x,t) \exp\left(\phi(\boldsymbol{\theta})\right) \right] \frac{\partial \phi(\boldsymbol{\theta})}{\partial \theta},$$

while its second derivative takes the form:

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \theta^2} = -\sum_{x,t} \left[E(x,t) \exp\left(\phi(\boldsymbol{\theta})\right) \right] \left(\frac{\partial \phi(\boldsymbol{\theta})}{\partial \theta} \right)^2,$$

as long as $\phi(\boldsymbol{\theta})$ is linear in the elements of $\boldsymbol{\theta}$. The age-period interactions through $\beta_x^{(2)}$ and $\kappa_t^{(2)}$ in the exponential term of the above function leads to non-linear maximisation techniques. Commonly, approximate estimates are obtained using Newton's method until some tolerance on subsequent iterations of equation (1.2.3) is satisfied. To achieve a unique set of parameters, the estimates are scaled according to the constraints given in equation (1.2.2). Generally, the use of different constraints would lead to different parameter estimates. Of course, the fitted rates and the associated forecasts should not differ under the various sets of constraints.

A pseudo-algorithm for the estimation scheme follows:

Algorithm 1a

- Initialise the parameter vector $\widehat{\boldsymbol{\theta}} = \left(\widehat{\beta}_x^{(1)}, \widehat{\beta}_x^{(2)}, \widehat{\kappa}_t^{(2)} \right)$, calculate $\ell_s = \ell\left(\widehat{\boldsymbol{\theta}}\right)$ and set the tolerance, ϵ.
- For $i = 1, \ldots, m$:
  - Update the $\beta_x^{(2)}$ set by the approximation:

$$\widehat{\beta}_{x_i}^{(2)} = \widehat{\beta}_{x_i}^{(2)} - \left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_{x_i}^{(2)}} \Big/ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\left(\partial \beta_{x_i}^{(2)}\right)^2} \right) \Bigg| \widehat{\boldsymbol{\theta}}.$$

  - Update the $\beta_x^{(1)}$ set by:

$$\widehat{\beta}_{x_i}^{(1)} = \log \left(\frac{\sum_t D(x_i,t)}{\sum_t E(x_i,t) \exp\left(\widehat{\beta}_{x_i}^{(2)} \widehat{\kappa}_t^{(2)} \right)} \right).$$

- For $i = 1, \ldots, n$:
  - Update the $\kappa_t^{(2)}$ set by the approximation:

$$\widehat{\kappa}_{t_i}^{(2)} = \widehat{\kappa}_{t_i}^{(2)} - \left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \kappa_{t_i}^{(2)}} \Big/ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\left(\partial \kappa_{t_i}^{(2)}\right)^2} \right) \Bigg| \widehat{\boldsymbol{\theta}}.$$

- Apply the constraints (1.2.2) by implementing the equations:

$$\widetilde{\beta}_x^{(1)} = \widehat{\beta}_x^{(1)} + \widehat{\beta}_x^{(2)} \left(\sum_t \widehat{\kappa}_t^{(2)} \Big/ n \right),$$
$$\widetilde{\kappa}_t^{(2)} = \sum_x \widehat{\beta}_x^{(2)} \times \left(\widehat{\kappa}_t^{(2)} - \sum_t \widehat{\kappa}_t^{(2)} \Big/ n \right),$$
$$\widetilde{\beta}_x^{(2)} = \widehat{\beta}_x^{(2)} \Big/ \left(\sum_x \widehat{\beta}_x^{(2)} \right).$$

- Set, $\widehat{\boldsymbol{\theta}} = \left(\widetilde{\beta}_x^{(1)}, \widetilde{\beta}_x^{(2)}, \widetilde{\kappa}_t^{(2)} \right)$, calculate $\ell_u = \ell\left(\widehat{\boldsymbol{\theta}}\right)$ and $\delta = \ell_s - \ell_u$; check whether $\delta < \epsilon$ and if so return the latest values, otherwise return to step 2,

where m and n are the number of ages and years, respectively, as mentioned in the Introduction. **Algorithm 1a** is applied to the EW and Canada data-sets. The parameter estimates for both datasets are shown in Figure 1.1. Since the $\beta_x^{(1)}$ set describes the general pattern of the log-mortality surface across the ages of the data-set, their trend is increasing for both cases, indicating that mortality becomes heavier as age increases. The set of $\beta_x^{(2)}$ are the terms indicating which ages have rapid or slow changes relative to changes in the period index. These are expected to be positive, since otherwise it is implied that mortality rises at those ages that it falls for others (Lee and Carter, 1992).

Under the original LC estimation, the age effects $\beta_x^{(2)}$ usually tend to lack smoothness and become increasingly jagged over time, which is non-intuitive and may be problematic in practical applications (Booth and Tickle, 2008). It is also known that lack of smoothness in $\beta_x^{(2)}$ results in the violation of the residuals' *i.i.d.* standard normality assumption for future forecasts (Cairns et al., 2008). The Poisson assumption employed here is intended to mitigate this effect under a more consistent statistical framework. Czado et al. (2005) used the Poisson log-bilinear model together with Bayesian estimation to impose smoothness. Delwarde et al. (2007) apply a P-splines methodology to the age effects to overcome this issue.

The period effects $\kappa_t^{(2)}$ develop an almost linear decreasing trend across time for both data-sets. Their evolution is more volatile for the EW dataset than for the Canada population. The linear decrease of the time series seems to be a shared characteristic across the majority of developed countries (Tuljapurkar et al., 2000). In most cases the assumption is not valid for the whole of the available observation period, but instead it seems that the steepness of $\kappa_t^{(2)}$ changes from time to time.

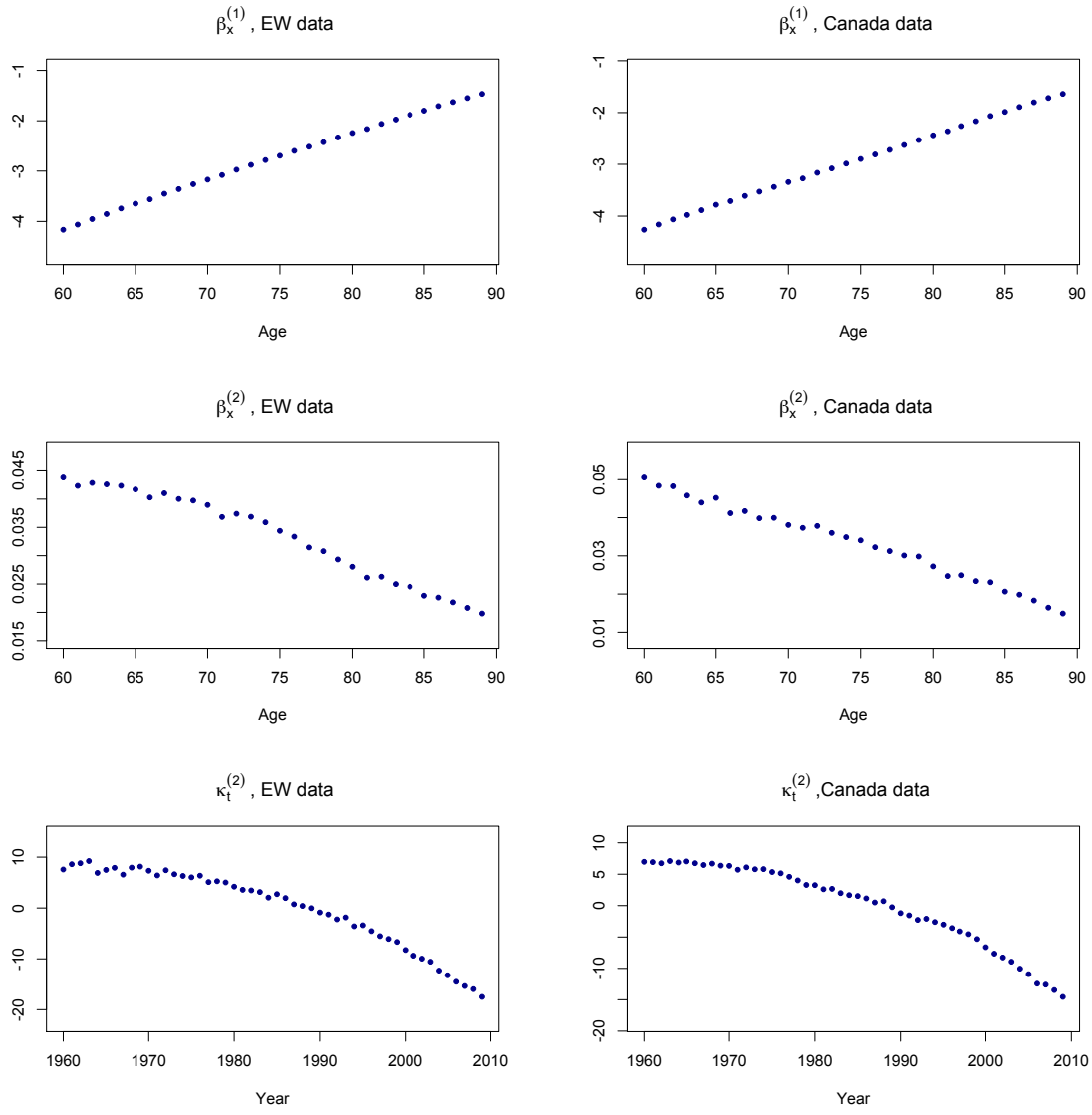

Figure 1.1: Series of parameter estimates, $\beta_x^{(1)}$, $\beta_x^{(2)}$ and $\kappa_t^{(2)}$, of the LC model, EW and Canada data.

Such a behaviour exhibits a similar evolution of mortality due to parallel socio-economic factors, likeness in lifestyle and correspondence of medical advancements. The constant decreasing pattern of the time series motivated the use of a random walk with drift model at the original development of the model (Lee and Carter, 1992). However, the chosen model is not prescribed for using the LC model, and indeed its misspecification is likely to lead in erratic forecasts, since it is the main extrapolative component of the model.Therefore, several alternatives have been suggested in cases that data had been found inconsistent with it. For example, Booth et al. (2002) attempt to maximise the fit of the overall model, by restricting the fitting period such as the linear decrease assertion for $\kappa_t^{(2)}$ is adequately supported for the Australian data-set. By fixing that deficiency for their specific data-set, they also achieve to meet better the invariant age component assumption of the model. Chan

et al. (2008) examine whether the period factor of the LC model are best described by deterministic broken trend models instead of the random walk. Li et al. (2011) also argue on the linearity of $\kappa_t^{(2)}$ and detect statistically significant structural breakpoints for the EW and USA populations.

1.2.3 Summary and Extensions

The LC model is highly important in its class. It can be seen as an important step in the introduction and wider acceptance of formal statistical methods to modelling mortality in a dynamic context. Moreover, it accommodates simplicity and transparency to explain mortality over the main part of the lifespan by also explaining a large amount of variability in the data.

Despite the model's significance, its drawbacks have served as starting points for various extensions and improvements. First, the model includes a single time varying factor. Even for a restricted subset of a whole life data-set, as for example here with ages from 60 to 89, mortality is likely to be driven by more than one factor. Hence the model is most probably insufficient in describing the mortality dynamics adequately. Second, the single factor results in mortality improvements at all ages being perfectly correlated. The incorporation of a non-trivial correlation structure might also be faced by the inclusion of additional stochastic factors in the model. For example, Renshaw and Haberman (2003b) proposed the model given as:

$$\log\big(m(x,t)\big) = \beta_x^{(1)} + \beta_x^{(2)}\kappa_t^{(2)} + \beta_x^{(3)}\kappa_t^{(3)}, \tag{1.2.4}$$

where $\beta_x^{(3)}$ and $\kappa_t^{(3)}$ are additional sets of age and period related parameters, respectively. In general, the inclusion of further interaction terms in the modelling equation of the original model determines a family of LC-type models. Such extensions have been considered, for example by Hyndman and Ullah (2007), who developed a functional multiple weighted principal components analysis method to determine the mortality basis functions. Directly extending the LC model, Booth et al. (2002) include up to five bilinear age-period terms, leading to higher order SVD. Although models developed under this approach solve the issue of perfectly correlated mortality improvements, they also yield increasingly complicated structures which are not transparent enough. Moreover, several such practical implementations lead to models that seriously lack robustness, for example fitting the model to a subset of the original fitting period yields qualitative different parameter sets, and therefore, the projections they produce cannot be reliable (Cairns et al., 2011a).

Another drawback of the LC model is connected with the interpretation of the interaction age-effects, $\beta_x^{(2)}$. That set of parameters determines the amount of associated

uncertainty with the future death rates for fixed age x. Essentially, the confidence bands of forecasted rates are proportionally wide to the $\beta_x^{(2)}$. In turn and given the results of the model as in Figure 1.1, predicted uncertainty in future death rates will be significantly lower at high ages. However, in reality, variations in mortality improvements at old ages have been significantly high, impairing the adequacy of the model in that sense (Cairns et al., 2008). The above points combined with the poor fit of the model, as will be shown in Section 1.4, introduce biased estimates which result in poor forecasting properties. The confidence bands are narrow and the estimation bias appears as an inherited characteristic of the model to the projections. Although more formal statistical frameworks attempt to face that issue, such as the Poisson model employed here, the standardised residuals under any set of assumptions appear to violate the *i.i.d.* requirement.

If the standardised residuals of deaths are plotted, as done in Section 1.4.1, there are identifiable patterns which indicate the lack of required additional effects in the model. The model generally fails to capture excessive volatile rate distortions at higher ages when these occur, and the systematic bias at lower ages generally results in biased forecasts of age-specific mortality rates (Lee and Miller, 2001). Additionally, the diagonal clusters dominating the plots of Figure 1.5 correspond to the apparent cohort effect of the data. Cohort effect might be defined as the dependence between mortality rates and the year of birth of the examined group of lives. The impact of the cohort effect has been identified in several data-sets, and its importance has been highlighted by various studies over time, *e.g.* in Richards et al. (2006). As already noted in the Introduction, in a data-set of x ages and t years, the cohort year of births of the lives, c, are given as:

$$c = t - x. \tag{1.2.5}$$

The generalised LC family of models might be written as:

$$\log\big(m(x,t)\big) = \sum_i \beta_x^{(i)} \kappa_t^{(i)} \gamma_c^{(i)}, \tag{1.2.6}$$

where the newly introduced $\gamma_c^{(i)}$ terms indicate cohort related effects. All the models noted so far can be based under the above formulation. For example, the LC model occurs for $i = 2$ and $\kappa_t^{(1)} = \gamma_c^{(i)} = 1$, over all periods t and cohort groups c. Similarly, the Renshaw-Haberman model, given in equation (1.2.4) arises for $i = 3$ and $\kappa_t^{(1)} = \gamma_c^{(i)} = 1$.

The initiation of models including cohort effects might be traced back to the field of medical statistics (Osmond, 1985). The addition of cohort effects to simpler modelling structures has gained popularity and it is almost the mainstream method of developing new mortality models nowadays. For example, Renshaw and Haberman (2006), sim-

ilarly to the model of equation 1.2.4, attempt the following cohort-enhanced version of the LC model:

$$\log\big(m(x,t)\big) = \beta_x^{(1)} + \beta_x^{(2)}\kappa_t^{(2)} + \beta_x^{(3)}\gamma_c^{(3)}. \tag{1.2.7}$$

In its initial form, the interaction effects, $\beta_x^{(2)}$ and $\beta_x^{(3)}$, are set equal to 1, and the implementation is done under their usual Poisson GLM framework, with the additional allowance for overdispersion of the underlying distribution. For this model and in contrast to model (1.2.4), we have an additional interaction between the third set of age effects, $\beta_x^{(3)}$, and the cohort effects, $\gamma_c^{(3)}$. However, the set of $\beta_x^{(1)}$ are still estimated according to the original LC method and thus, are set to be equal to $\frac{1}{n}\sum_t \log\big(\widetilde{m}(x,t)\big)$ and are treated as constants throughout the iteration scheme. Renshaw and Haberman (2006) also analyse other possibilities for the interaction terms, namely $\beta_x^{(2)} = 1$ and $\beta_x^{(3)}$ parameters to be estimated, and vice versa. Cairns et al. (2009) calibrate the model using fully iterative MLE under the Poisson assumption for deaths, while all age effects are treated as unknown parameter sequences which need to be estimated. In contrast to Renshaw and Haberman (2006), they include the series of $\beta_x^{(1)}$ in their iterative scheme too.

In principle, models including age, period and cohort effects are well–known for their identifiability problems due to the linear dependence of their parameters, as evident from equation (1.2.5). Essentially, each cohort effect might approximately be replaced by appropriately adjusted age and period effects. That affects the robustness of the parameter estimates and in turn, the reliability of the forecasts. For the Poisson implementation of the model given in equation (1.2.7), Cairns et al. (2009) found that changing the fitting period of the estimation results in qualitatively different parameter sets. They conclude that the resulting likelihood has more than one maximum. Changing the fitting period or the range of ages results in the optimiser achieving a qualitatively different local maximum. On the contrary, a robust model should possess a likelihood which has a unique and unchanging global maximum, attained by fitting the model for any period or range of years (Cairns et al., 2008). Although that generalised approach improves the performance of the LC model, excessive differences in parameters between subsets of the same dataset is not perceived as a desirable feature of a model.

A simpler version of the cohort-enhanced LC model of equation (1.2.7) is taken by substituting the age effects $\beta_x^{(2)}$ and $\beta_x^{(3)}$ by the constant $1/m$. In this case we get the simple Age-Period-Cohort (APC) model given as:

$$\log\big(m(x,t)\big) = \beta_x^{(1)} + \frac{1}{m}\kappa_t^{(2)} + \frac{1}{m}\gamma_c^{(3)}. \tag{1.2.8}$$

Under this parameterisation, the model has been examined by Currie (2006), who used P-splines to smooth out the age parameters, as well as by Cairns et al. (2009) under no such smoothness restriction. We use the model as a representative of the LC family of models including cohort effects throughout the Thesis. It is implemented under the Poisson model, and the unique maximum of the likelihood is obtained by applying the constraints:

$$\sum_t \kappa_t^{(2)} = 0, \sum_c \gamma_c^{(3)} = 0, \sum_c c\gamma_c^{(3)} = 0. \tag{1.2.9}$$

The estimation framework introduced in section 1.2.2 may still be used through the function of parameters:

$$\phi(\boldsymbol{\theta}) = \log\left(m(x,t;\boldsymbol{\theta})\right) = \beta_x^{(1)} + \frac{1}{m}\kappa_t^{(2)} + \frac{1}{m}\gamma_c^{(3)},$$

where $\boldsymbol{\theta}$ now includes the sets of $\beta_x^{(1)}, \kappa_t^{(2)}$ and $\gamma_c^{(3)}$. The algorithm is simplified due to the lack of interactions of latent parameters, so that Newton's method is no longer required to give an approximate solution.

A pseudo-algorithm for the estimation scheme follows:

Algorithm 1b

- Initialise the parameter vector $\widehat{\boldsymbol{\theta}} = \left(\widehat{\beta}_x^{(1)}, \widehat{\kappa}_x^{(2)}, \widehat{\gamma}_c^{(3)}\right)$, calculate $\ell_s = \ell\left(\widehat{\boldsymbol{\theta}}\right)$ and set the tolerance, ϵ.
- For $i = 1, \ldots, m$:
  - Update the $\beta_x^{(1)}$ set by:

$$\widehat{\beta}_{x_i}^{(1)} = \log\left(\frac{\sum_t D(x_i,t)}{\sum_t E(x_i,t)\exp\left(\frac{\widehat{\kappa}_t^{(2)}}{m} + \frac{\widehat{\gamma}_{t-x_i}^{(3)}}{m}\right)}\right).$$

- For $i = 1, \ldots, n$:
  - Update the $\kappa_t^{(2)}$ set by:

$$\widehat{\kappa}_{t_i}^{(2)} = m\log\left(\frac{\sum_x D(x,t_i)}{\sum_x E(x,t_i)\exp\left(\widehat{\beta}_x^{(1)} + \frac{\widehat{\gamma}_{t_i-x}^{(3)}}{m}\right)}\right).$$

- For $i = 1, \ldots, c$:

– Update the $\gamma_c^{(3)}$ set by:

$$\widehat{\gamma}_{c_i}^{(3)} = m \log \left(\frac{\sum\limits_{c_i=t-x} D(x,t)}{\sum\limits_{c_i=t-x} E(x,t) \exp\left(\widehat{\beta}_x^{(1)} + \dfrac{\widehat{\kappa}_t^{(2)}}{m}\right)} \right).$$

- Apply the constraints (1.2.9) by implementing the equations:

$$\widetilde{\beta}_x^{(1)} = \widehat{\beta}_x^{(1)} + \left(\frac{\sum\limits_t \widehat{\kappa}_t^{(2)}}{n} + \frac{\sum\limits_c \widehat{\gamma}_c^{(3)}}{(n+m-1)} - \frac{\sum\limits_c (c-\bar{c})\widehat{\gamma}_c^{(3)}}{\sum\limits_c (c-\bar{c})^2} \right) \Big/ m,$$

$$\widetilde{\kappa}_t^{(2)} = \widehat{\kappa}_t^{(2)} - \sum_t \widehat{\kappa}_t^{(2)} \Big/ n + \sum_c \widehat{\gamma}_c^{(3)} \Big/ (n+m-1),$$

$$\widetilde{\gamma}_c^{(3)} = \widehat{\gamma}_c^{(3)} - \sum_c \widehat{\gamma}_c^{(3)} \Big/ (n+m-1) - (c-\bar{c}) \left(\sum_c (c-\bar{c})\widehat{\gamma}_c^{(3)} \Big/ \sum_c (c-\bar{c})^2 \right).$$

- Set, $\widehat{\boldsymbol{\theta}} = \left(\widetilde{\beta}_x^{(1)}, \widetilde{\kappa}_t^{(2)}, \widetilde{\gamma}_c^{(3)}\right)$, calculate $\ell_u = \ell\left(\widehat{\boldsymbol{\theta}}\right)$ and $\delta = \ell_s - \ell_u$; check whether $\delta < \epsilon$ and if so return the latest values, otherwise return to step 2.

The symbol $\bar{c}$ has the meaning of the mean year of birth across all cohorts involved in the data-set. The algorithm is slightly complicated by the summations which require specific entries of the individual parameter vectors. However, this is easily solved by appropriate indexing.

Algorithm 1b is applied to the EW and Canada data-sets, and the resulting parameter estimates are shown in Figure 1.2. The age effects $\beta_x^{(1)}$ maintain the form they had under the LC model, indicating increasing mortality with age. The decreasing pattern of the period factors $\kappa_t^{(2)}$ postulate the improvement of mortality throughout time. Notably, and in line with the results of the LC model, the trend is smoother for the Canadian dataset compared to the estimates for the EW dataset.

The plots of the cohort effects omit estimates involved in less than five cells within the dataset, as well as the effects of the cohorts with year of birth 1918,1919 and 1920 for the EW dataset. On the one hand, the former set of estimates are generally not regarded reliable due to the insufficient information for their determination. On the other hand, the latter estimates for the cohort effects of the EW data are considered excessive outliers applying for certain years of births though time and distort the parameter estimates of the underlying stochastic model for $\gamma_c^{(3)}$ (Cairns et al., 2009). The dominant explanation is due to a combination of the Spanish flu and effects linked to the end of World War I. Additionally, the cohort 1919 has been suspicious for underlying exposures misspecification, which motivates for modelling their uncertainty jointly within a mortality model (Richards, 2008, Cairns et al., 2014). Lastly,

note the rapid decline of the cohort effects after the mid-1920's. It has been generally suggested that a possible reason for the rapid mortality improvement experienced by the generations after that time is less to do with beneficial conditions applying to those generations, and much more to do with the adverse conditions applying to the previous generations (Willets, 2004).

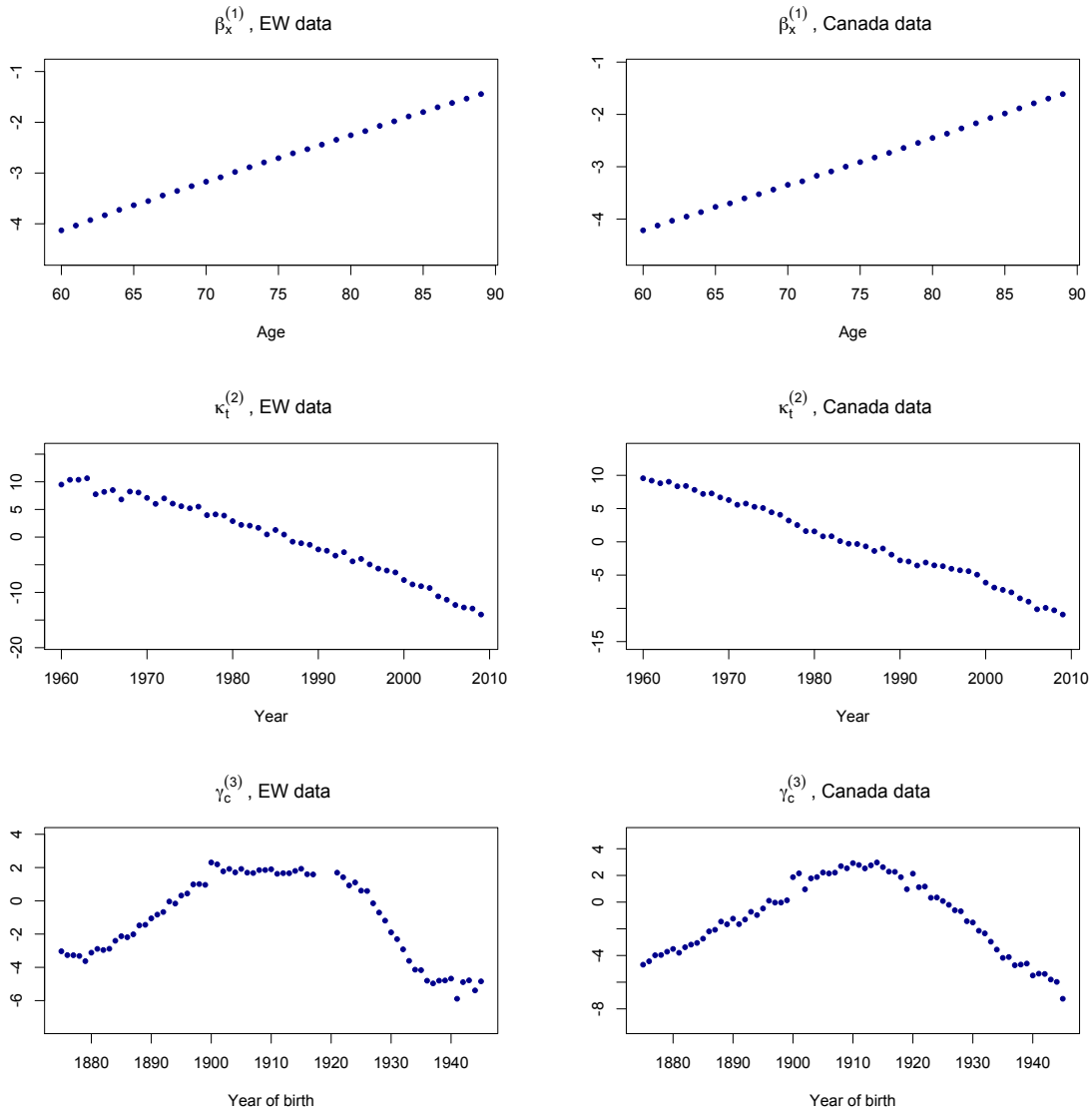

Figure 1.2: Series of parameter estimates, $\beta_x^{(1)}$, $\kappa_t^{(2)}$ and $\gamma_c^{(3)}$, of the APC model, EW and Canada data.

1.3 The Cairns-Blake-Dowd Model

The Cairns-Blake-Dowd (CBD) model (Cairns et al., 2006a) is the second benchmark time-series mortality model use in the Thesis. Its design is motivated from forecasting mortality rates directly and it arises as the stochastic counterpart of the family of

models introduced in Perks (1932). Although it is derived from different principles than the LC model, it is similarly simple. The CBD model utilises the observation that for older age groups the logistic transformation of $q(x,t)$ seems to be reasonably linear with age. That observation is commonly used when modelling mortality for higher age bands and therefore, the model is designed to be used for those.

1.3.1 Description

In contrast to the LC model, its dynamics are driven by a bivariate process, so that it allows for non-trivially correlated changes in mortality across ages. Furthermore, the model assumes constant functional age effects, which allow for smoothness in mortality rates at adjacent ages, in contrast to the original LC model, where smoothness is not assured in any way. Lastly, different versions of the model may arise for different assignments of the age effects, e.g. (Cairns et al., 2006a, 2009).

The model as presented in its later version (Cairns et al., 2009), is given by:

$$\operatorname{logit}\big(q(x,t)\big) = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}),$$

where $\bar{x}$ is the mean of the range of ages of the dataset.

The functional age effects, namely $\beta_x^{(1)} = 1$ and $\beta_x^{(2)} = x - \bar{x}$, constitute a basis for the logit of the mortality curve and impose smoothness in mortality rates at adjacent ages. The two factors of the model are able to capture simultaneous divergent dynamics because of the same effect, whereas this is naturally impossible with the single factor of the LC model. The first period effects, $\kappa_t^{(1)}$, affect mortality at all ages in an equal way and they are thus interpreted as the level of the logit transformed mortality curve. The second period effects, $\kappa_t^{(2)}$, are the slope of the logit mortality curve.

The model has no identifiability problems since it is a full-rank regression model. This feature makes it highly important, since the lack of constraint requirements results in a *new-data invariant* model. This means that adding new years of observations in a given dataset will not alter the already estimated parameters under a maximum likelihood scenario. The importance of that property is also emphasised in Chan et al. (2014), where the dynamics underlying the CBD model are suggested as appropriate mortality indices and where more complex stochastic models are employed for their specification.

1.3.2 Estimation

Since it is a full-rank regression model the least squares (LS) estimates might be taken imminently. For example, for fixed time $t = t_j$ the LS estimators of the two period

effects $\kappa_{t_j}^{(1)}$ and $\kappa_{t_j}^{(2)}$ are:

$$\widehat{\kappa}_{t_j}^{(1)} = \frac{1}{m}\sum_{x=x_1}^{x_m} \text{logit}\big(\widetilde{q}(x,t_j)\big) \text{ and } \widehat{\kappa}_{t_j}^{(2)} = \frac{\sum_{x=x_1}^{x_m} \text{logit}\big(\widetilde{q}(x,t_j)\big)\big(x-\bar{x}\big)}{\sum_{x=x_1}^{x_m}\big(x-\bar{x}\big)^2},$$

under the convention that $\widetilde{q}(x,t)$ denote observed quantities. Working iteratively through all years of the dataset yields the full set of estimates for all period effects, $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$. The estimator of $\kappa_t^{(1)}$'s indicates that they depict the average profile of the logit-mortality surface across the ages of the dataset. The homoscedastic errors implied by the LS solution is questionable, as in the case of the LC model. An alternative is the weighted LS approach, so that the different variabilities are taken into account for different cells of the population, as detailed when the LC model was examined.

Estimating the model under the Poisson assumption for deaths allows for a unified comparison of the stochastic mortality models in this Thesis under the same source of natural risk. In this case, the parameter vector of the model, $\boldsymbol{\theta}$, consists of the series of period effects, $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$. Under the constant force of mortality assumption, the likelihood of the Poisson model is used by connecting $m(x,t)$ and $q(x,t)$ through equation (1.1.7). Therefore, similarly to the LC estimation scheme, let:

$$\psi(\boldsymbol{\theta}) = \text{logit}\big(q(x,t;\boldsymbol{\theta})\big) = \kappa_t^{(1)} + \kappa_t^{(2)}\big(x-\bar{x}\big). \tag{1.3.1}$$

Then, note that under the constant force of mortality assumption equation (1.3.1) is written:

$$m(x,t;\boldsymbol{\theta}) = \log\Big[1+\exp\big(\psi(\boldsymbol{\theta})\big)\Big],$$

and the log-likelihood takes the form:

$$\ell(\boldsymbol{\theta}) = \sum_{x,t}\left\{D(x,t)\log\left\{\log\Big[1+\exp\big(\psi(\boldsymbol{\theta})\big)\Big]\right\} - E(x,t)\log\Big[1+\exp\big(\psi(\boldsymbol{\theta})\big)\Big]\right\} + C,$$

where C is some constant independent of the parameters of the model. The objective function is non-linear in the parameters of the model. The non-linear relationship yields slightly complicated, but nevertheless easily computable, first and second partial derivatives *w.r.t.* the period effects $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$. For a fixed entry of the parameter vector, $\boldsymbol{\theta}$, we get:

$$\frac{\partial\ell(\boldsymbol{\theta})}{\partial\theta} = \sum_{x,t}\left\{\left[\frac{D(x,t)}{m(x,t;\boldsymbol{\theta})} - E(x,t)\right]q(x,t;\boldsymbol{\theta})\frac{\partial\psi(\boldsymbol{\theta})}{\partial\theta}\right\},$$

and

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \theta^2} = \sum_{x,t} \left\{ \left[D(x,t) \left(p(x,t;\boldsymbol{\theta}) - \frac{q(x,t;\boldsymbol{\theta})}{m(x,t;\boldsymbol{\theta})} \right) - E(x,t) \frac{p(x,t;\boldsymbol{\theta})}{m(x,t;\boldsymbol{\theta})} \right] \left(\frac{\partial \psi(\boldsymbol{\theta})}{\partial \theta} \right)^2 \right\}.$$

Although not as obvious as in the LC case, the second derivatives *w.r.t.* each component of the parameter vector are still negative. The parameter estimates are obtained through an iterative application of Newton's method until some tolerance in successive iterations of the objective function is satisfied.

A pseudo-algorithm for the estimation scheme follows:

Algorithm 2a

- Initialise the parameter vector $\widehat{\boldsymbol{\theta}} = \left(\widehat{\kappa}_t^{(1)}, \widehat{\kappa}_t^{(2)} \right)$, calculate $\ell_s = \ell\left(\widehat{\boldsymbol{\theta}}\right)$ and set the tolerance, ϵ.
- For $i = 1, \ldots, n$:
  - Update the $\kappa_t^{(1)}$ set by the approximation:

$$\widehat{\kappa}_{t_i}^{(1)} = \widehat{\kappa}_{t_i}^{(1)} - \left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \kappa_{t_i}^{(1)}} \Big/ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\left(\partial \kappa_{t_i}^{(1)} \right)^2} \right) \Bigg| \widehat{\boldsymbol{\theta}}.$$

  - Update the $\kappa_t^{(2)}$ set by the approximation:

$$\widehat{\kappa}_{t_i}^{(2)} = \widehat{\kappa}_{t_i}^{(2)} - \left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \kappa_{t_i}^{(2)}} \Big/ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\left(\partial \kappa_{t_i}^{(2)} \right)^2} \right) \Bigg| \widehat{\boldsymbol{\theta}}.$$

- Calculate $\ell_u = \ell\left(\widehat{\boldsymbol{\theta}}\right)$ and $\delta = \ell_s - \ell_u$; check whether $\delta < \epsilon$ and if so return the latest values, otherwise return to step 2.

The fact that the CBD model directly refers to the mortality rates also allows the use of the Binomial model, as given in Section 1.1.4. The adoption of any of these assumptions is clearly a choice of implementation and the actual differences in practice are negligible.

Algorithm 2a is applied to the EW and Canada data-sets. The estimates are shown in Figure 1.3. Since the $\kappa_t^{(1)}$'s describe the way mortality is affected through time, they develop the decreasing trend which indicates the observed mortality improvement across all ages. On the other hand, the increase in the $\kappa_t^{(2)}$'s series indicates increase in the steepness of the logit-transformed mortality curve, which in turn means that mortality improves faster for younger ages than for older. The $\kappa_t^{(1)}$ estimates are again smoother for the Canada population. Also, in contrast to the awkward increase of

the $\kappa_t^{(2)}$'s for the final estimation years of the EW data-set, that period seems rather flat for Canada. However, generally the plots of the parameter estimates show that the dynamics are very similar, indicating the analogous trend of mortality between the two populations.

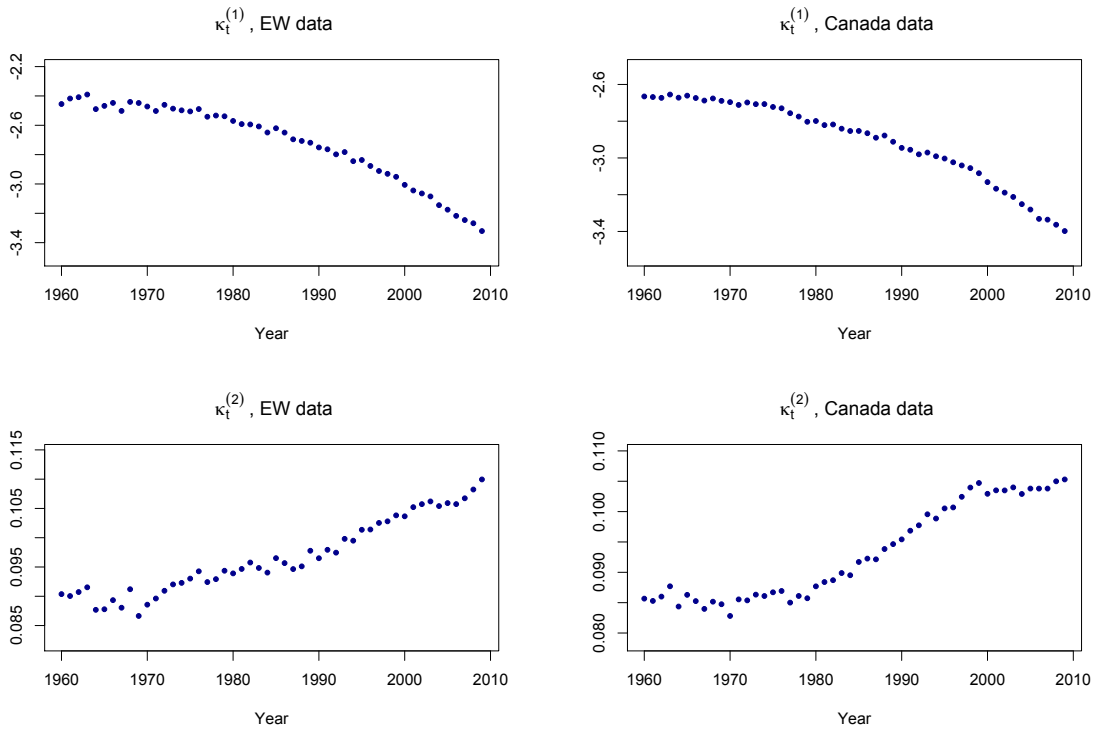

Figure 1.3: Series of parameter estimates, $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$, of the CBD model, EW and Canada data.

1.3.3 Summary and Extensions

The two period effects of the CBD model allow for further flexibility, so that more sophisticated mortality improvements are interpreted and incorporated in the forecasts. The dynamics of the model have been examined under alternative models beyond the random walk. For example, Sweeting (2011) fits a trend-change model to the period effects, after he estimates them for all available years in the dataset which consists of years from 1841 to 2005. It is noted that in the long term the period factors exhibit variation around a trend rather than following random walks. Additionally, the trend changes are observed to happen in the same instances for the two sets of period parameters, while there is strong negative correlation between the direction of them. Cannon (2009) compares the random walk assumption versus two alternative parameterisations. In the first case there is a single stochastic trend and the two time series are co-integrated and in the second, both effects are driven around deterministic, rather than stochastic, trends.

However, examination of the goodness of fit of the model shows that it does not fit the data sufficiently well, and that the underlying modelling assumptions are violated. The main problem behind those issues is that the model does not include the required cohort effects. Therefore, the model may be taken as a basis of a family of related models so that it is extended by the inclusion of additional factors. Along similar lines to the generalisation of the LC model, the extended form of the corresponding CBD family of models is written as:

$$\text{logit}(q(x,t)) = \sum_i \beta_x^{(i)} \kappa_t^{(i)} \gamma_{t-x}^{(i)}, \qquad i = 1, 2, \ldots$$

Cairns et al. (2009) consider extensions of the CBD model based on the general form given before. One of these extensions is described by the following equation:

$$\text{logit}\big(q(x,t)\big) = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}) + \kappa_t^{(3)}((x - \bar{x})^2 - \sigma_x^2) + \gamma_c^{(4)},$$

where $\bar{x}$ is as in the original CBD model and σ_x^2 the corresponding variance of the ages of the data-set in hand. Beyond the inclusion of the factor $\gamma_c^{(4)}$ to capture the cohort effect, there is an additional age-period interaction term. Due to being quadratic in age form, it represents the curvature of the logit-rates, and has been found to provide statistically significant improvement. According to the order presented by Cairns et al. (2009) the model is termed M7, which is also adopted here.

The likelihood and the corresponding derivatives developed in section 1.3.2 might still be used under the following parameterisation:

$$\psi(\boldsymbol{\theta}) = \text{logit}\big(q(x,t;\boldsymbol{\theta})\big) = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}) + \kappa_t^{(3)}((x - \bar{x})^2 - \sigma_x^2) + \gamma_c^{(4)},$$

where the parameter vector $\boldsymbol{\theta}$ includes the period effects, $\kappa_t^{(1)}$, $\kappa_t^{(2)}$ and $\kappa_t^{(3)}$, and the cohort factors, $\gamma_c^{(4)}$. The unique maximum of the parameter space is determined by the set of constraints:

$$\sum_c \gamma_c^{(4)} = 0, \sum_c c\gamma_c^{(4)} = 0, \sum_c c^2 \gamma_c^{(4)} = 0. \tag{1.3.2}$$

Although a simpler alternative such as $\sum_t \kappa_t^{(i)} = 0$, $i = 1, 2, 3$ would be simpler, the idea behind that choice stems from the requirement for $\gamma_c^{(4)}$ to capture the pure cohort effect without tending to adjust due to lack of additional age-period effects. Effectively, based on expressing $\gamma_c^{(4)}$ as

$$\gamma_c^{(4)} = \phi_1 + \phi_2 c + \phi_3 c^2 + \epsilon_c, \tag{1.3.3}$$

where $\epsilon_c \sim N(0, \sigma_\epsilon^2)$ *i.i.d.* for all involved cohort years of birth, c, the least squares solution, $\left(\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3\right)$, is used to map $\gamma_c^{(4)}$ to ϵ_c, and by corresponding adjustments to $\left(\kappa_t^{(1)}, \kappa_t^{(3)}, \kappa_t^{(3)}\right)$ satisfy equations (1.3.2). Hence, the fitted $\gamma_c^{(4)}$ will fluctuate around 0, so that a mean-reverting process might be facilitated to model them. An alternative choice for establishing identifiability for the same model may be found in Haberman and Renshaw (2011).

A pseudo-algorithm for the estimation scheme follows:

Algorithm 2b

- Initialise the parameter vector $\widehat{\boldsymbol{\theta}} = \left(\widehat{\kappa}_t^{(1)}, \widehat{\kappa}_t^{(2)}, \widehat{\kappa}_t^{(3)}, \widehat{\gamma}_c^{(4)}\right)$, calculate $\ell_s = \ell\left(\widehat{\boldsymbol{\theta}}\right)$ and set the tolerance, ϵ.

- For $i = 1, \ldots, n$ and $j = 1, 2, 3$:

  - Update the $\kappa_t^{(j)}$ set by the approximation:

$$\widehat{\kappa}_{t_i}^{(j)} = \widehat{\kappa}_{t_i}^{(j)} - \left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \kappa_{t_i}^{(j)}} \Big/ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\left(\partial \kappa_{t_i}^{(j)}\right)^2} \right) \Bigg| \widehat{\boldsymbol{\theta}}.$$

- For $i = 1, \ldots, c$:

  - Update the $\gamma_c^{(4)}$ set by:

$$\widehat{\gamma}_{c_i}^{(4)} = \widehat{\gamma}_{c_i}^{(4)} - \left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \gamma_{t_i}^{(4)}} \Big/ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\left(\partial \gamma_{c_i}^{(4)}\right)^2} \right) \Bigg| \widehat{\boldsymbol{\theta}}.$$

- Apply the constraints given in equations (1.3.2) first by obtaining the solution $\left(\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3\right)$ of regression (1.3.3), and then by implementing:

$$\begin{aligned}
\widetilde{\gamma}_c^{(4)} &= \widehat{\gamma}_c^{(4)} - \hat{\phi}_1 - \hat{\phi}_2 c - \hat{\phi}_3 c^2 \\
\widetilde{\kappa}_t^{(1)} &= \widehat{\kappa}_t^{(1)} + \hat{\phi}_1 + \hat{\phi}_2 (t - \bar{x}) + \hat{\phi}_3 \Big((t - \bar{x})^2 + \tfrac{1}{m} \textstyle\sum_{i=x_1}^{x_m} (i - \hat{x})^2 \Big) \\
\widetilde{\kappa}_t^{(2)} &= \widehat{\kappa}_t^{(2)} - \hat{\phi}_2 - 2\hat{\phi}_3 (t - \bar{x}) \\
\widetilde{\kappa}_t^{(3)} &= \widehat{\kappa}_t^{(3)} + \hat{\phi}_3,
\end{aligned}$$

  to obtained the rebalanced set of estimates, $\left(\widetilde{\kappa}_t^{(1)}, \widetilde{\kappa}_t^{(2)}, \widetilde{\kappa}_t^{(3)}, \widetilde{\gamma}_c^{(4)}\right)$.

- Set $\widehat{\boldsymbol{\theta}} = \left(\widetilde{\kappa}_t^{(1)}, \widetilde{\kappa}_t^{(2)}, \widetilde{\kappa}_t^{(3)}, \widetilde{\gamma}_c^{(4)}\right)$, calculate $\ell_u = \ell\left(\widehat{\boldsymbol{\theta}}\right)$ and $\delta = \ell_s - \ell_u$; check whether $\delta < \epsilon$ and if so return the latest values, otherwise return to step 2.

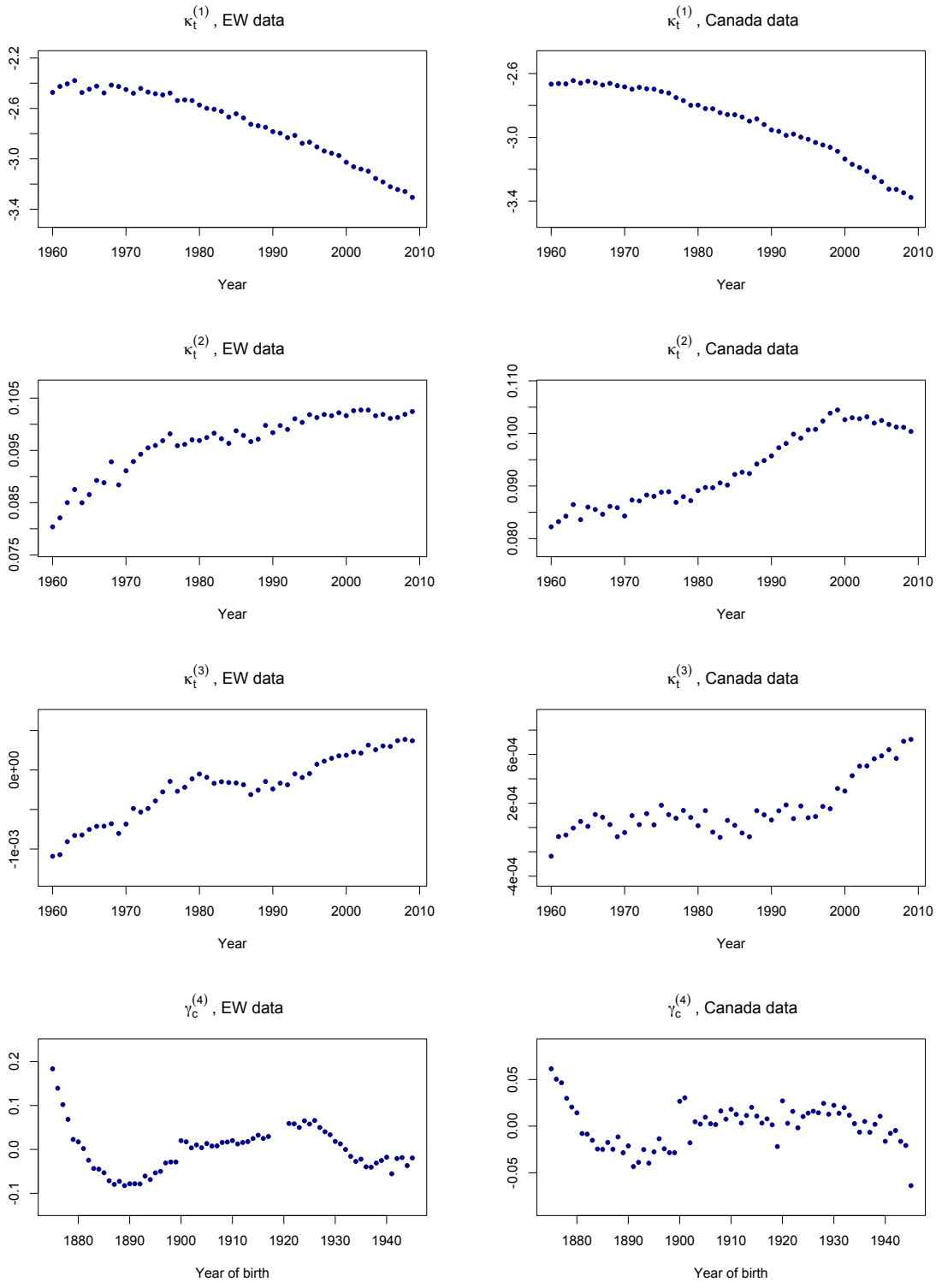

Figure 1.4: Series of parameter estimates, $\kappa_t^{(1)}$, $\kappa_t^{(2)}$, $\kappa_t^{(3)}$ and $\gamma_c^{(4)}$, of model M7, EW and Canada data.

Figure 1.4 shows the parameter estimates of model M7 for the EW and Canada datasets. The $\kappa_t^{(1)}$ period effects are very similar for both populations, although again smoother for the Canada population. Next, the set of $\kappa_t^{(2)}$ is increasing in a concave fashion for the EW dataset, in contrast to the clearly convex increase, at least until

2000, for the Canadian population.

However, the increasing trend in both cases makes the comments noted in Section 1.3.2 about them still valid. The set of $\kappa_t^{(3)}$ period effects is much more volatile for the Canada population compared to EW. The comments about the shape of increase made just before apply here as well. Lastly, the cohort factors $\gamma_c^{(4)}$ are by construction reverting around zero, and again appear more erratic in the case of the Canada data.

1.4 Comparisons and Forecasts

In order to distinguish among and compare competing mortality models, there are several selection and comparison criteria available. Examples of such selection criteria are the production of biologically reasonable long-term dynamics, the plausibility and consistency of central trajectories and uncertainty levels with historical trends and the ability of the model to generate sample paths for the calculation of prediction intervals. Comparison criteria might be divided into qualitative and quantitative. An example of a qualitative comparison criterion is the robustness of parameter estimates when, for example, a different range of years of the dataset is used to fit the model. Lastly, examples of quantitative comparison criteria are mostly derived from the goodness-of-fit model, either through the examination of the resulting residuals or through information criteria, such as the BIC.

The imminent practical application of mortality models, such as those developed in the present Chapter, is the projection of future mortality rates. As detailed in Section 1.2.1 mortality modelling is a two step procedure. Firstly, the model is fitted to the data so that the latent states of the various effects are estimated. Based on these, the parameters of the underlying stochastic processes for the random components are obtained. Finally, the models are projected to get the future rates. The dynamics of each stochastic mortality model are determined by the chosen processes for its random factors. Given the discrete time framework implied by the data, commonly ARIMA models have been used to describe these dynamics. Subsequently, the future rates might be used to derive further mortality related metrics, such as those presented in Section 1.1.5.

1.4.1 Comparison of Mortality Models

The assessment of the goodness-of-fit involves the examination of the residuals a mortality model produces. If $\widehat{m}(x,t)$ are the fitted rates of the model and the crude death rates $\widetilde{m}(x,t)$ are taken usual, under the Poisson assumption the deaths should be approximately normally distributed with mean and variance equal to $E(x,t)\widehat{m}(x,t)$

in each cell. If the model fits the data sufficiently well, the standardised residuals

$$\varepsilon(x,t) = \frac{\widetilde{m}(x,t) - \widehat{m}(x,t)}{\sqrt{\widetilde{m}(x,t)/E(x,t)}}, \tag{1.4.1}$$

are predicted to be approximately standard normally distributed both when one follows a certain age from one year to another and when death rates are compared over the range of ages for any fixed year (Dowd et al., 2010b).

Figure 1.5 shows positive and negative values of the standardised residuals of deaths from the application of the models presented in the previous sections in red and black squares, respectively. The persistent diagonal patterns observed in the LC and CBD plots violate the *i.i.d* assumption and indicate the lack of cohort effects in the models, or equivalently the existence of the cohort impact in the examined datasets. The observation is generally more intense when examining the EW population. This might also be deduced by the comparison of the magnitude of the cohort related parameters, $\gamma_c^{(4)}$, of model M7 in Figure 1.4 between the two datasets. The APC model also produces unsatisfactory plots. The main reason behind this behaviour is that regardless of the model containing cohort factors, the use of the constant age effects, that is $1/m$, is unable to capture the realised age-period and age-cohort interactions. Finally, model M7 returns the most adequate graphs. The model beyond its cohort factors also incorporates the curvature coefficients, $\kappa_t^{(3)}$, which have been found to improve significantly its statistical fit (Cairns et al., 2009).

Beyond the detected patterns, the matrices of residuals also exhibit excessively large values. Table 1.1 summarises the sample mean and sample variance of the estimated residuals shown in Figure 1.5. For the EW data-set, the LC and CBD models yield extremely high sample variance values, which appear to reduce in the case of the cohort-enhanced versions of the models. However, the magnitude is still much higher than acceptable levels, both for the APC and for model M7.

	EW Dataset		Canada Dataset	
Model	Mean	Variance	Mean	Variance
LC	9.724×10^{-3}	6.231	4.121×10^{-2}	2.684
APC	-1.342×10^{-1}	3.709	-9.794×10^{-2}	2.187
CBD	-1.184×10^{-1}	6.875	8.453×10^{-2}	2.113
M7	4.781×10^{-2}	2.281	2.074×10^{-3}	1.273

Table 1.1: Summary statistics of estimated standardised residuals of deaths for models LC, APC, CBD and M7, EW and Canada data.

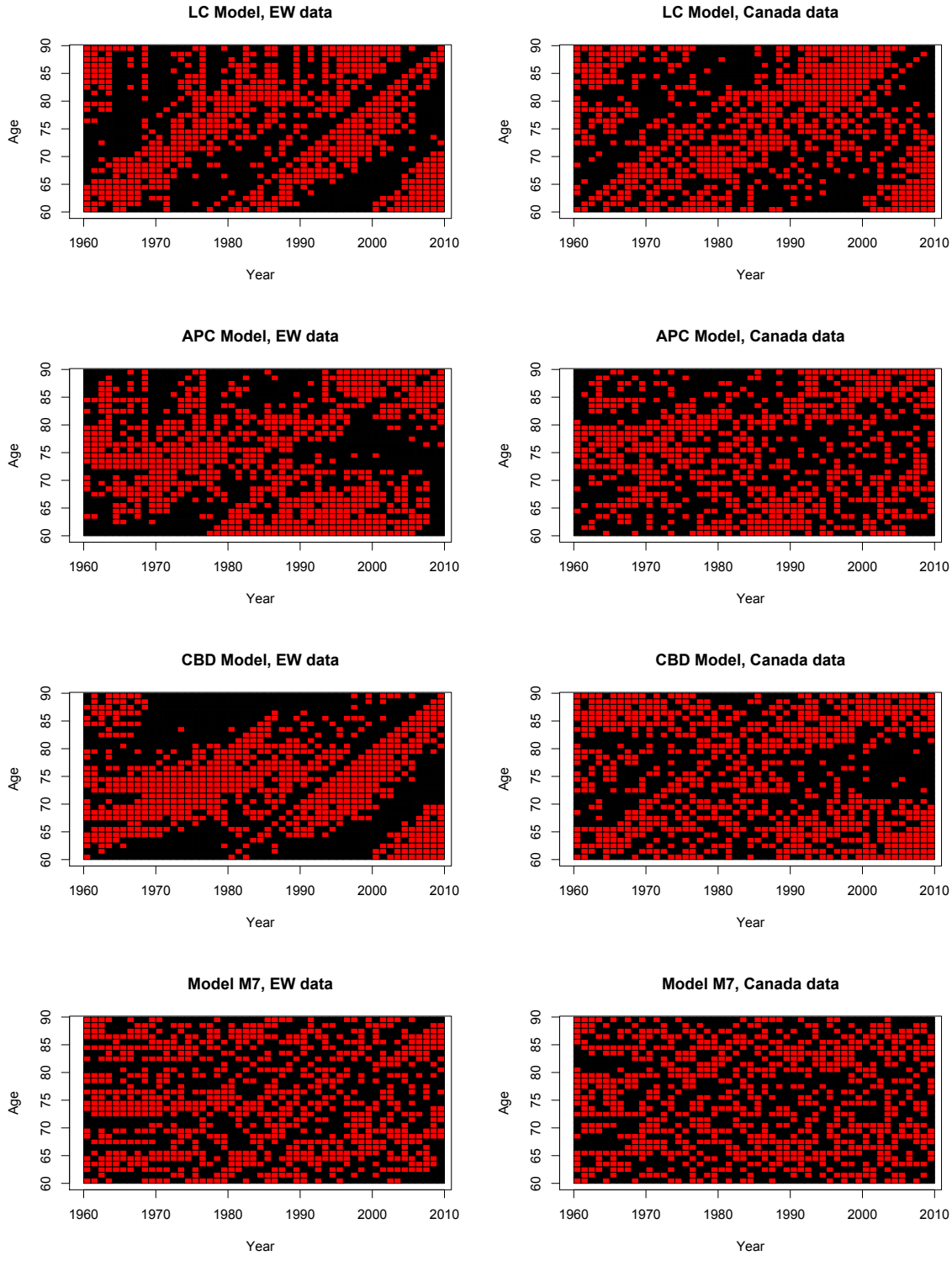

Figure 1.5: Standardised residuals of deaths from the LC, APC, CBD and M7 models, EW and Canada data. Red and black squares indicate positive and negative values, respectively.

The sample means, though, seem to be more consistent with the underlying zero-mean hypothesis. The models appear to fit better to the Canadian population rather than to that of EW, since in the first case all examined statistics are closer to the

modelling assumptions. For both data-sets, the most satisfactory performance arises under model M7.

Other quantitative means of comparison between competing models are based on the statistical properties of the suggested structures. The Bayesian Information Criterion (BIC) provides a tool for a balanced choice between quality of fit and parsimony. Beyond its ability to distinguish between models that are not nested, or even at all related, it penalises so that only significant improvements in the likelihood are awarded.

For a model $\mathcal{M}$ with corresponding log-likelihood $\ell_{\mathcal{M}}(\boldsymbol{\theta})$ and parameter estimates $\widehat{\boldsymbol{\theta}}$, if the sample size is N and the model has $\nu_{\mathcal{M}}$ parameters, then the BIC is given as:

$$BIC_{\mathcal{M}} = \ell_{\mathcal{M}}(\widehat{\theta}) - \frac{1}{2}\nu_{\mathcal{M}} \log(N).$$

The resulting BIC's and the respective ranking of the four mortality models is shown in Table 1.2. In terms of BIC, the CBD model performs much better in the case of the Canada data, in contrast to the LC model, for which the converse is true. Notably, the APC model ranks first in the case of the Canada population, although the results from the analysis of its standardised residuals would not suggest so. Model M7 clearly outperforms its competitors in the case of the EW data-set. The differences between the four models appear to be smaller for the Canada data. In summary, evidently some of the models fit more adequately the given data-sets than others. However, there cannot be drawn a definite conclusion on which model is clearly the ideal. Rather, there are models that are more appropriate than others for specific data-sets.

Model	BIC (rank), EW.	BIC (rank), Canada.
LC	$-12,774.91(3)$	$-9,252.12$ (4)
APC	$-10,317.01(2)$	$\mathbf{-8,704.19(1)}$
CBD	$-13,306.11(4)$	$-8,892.20$ (3)
M7	$\mathbf{-9,742.42(1)}$	$-8,767.34$ (2)

Table 1.2: BIC criterion and ranking amongst models LC, APC, CBD and M7, EW and Canada data.

There also exist alternative means of comparison beyond those considered in this Section. In the case of nested models, for example the CBD and M7, the likelihood ratio test might be used to test the hypothesis that the more general model is more appropriate. Such tests for the models considered here tend always to favour the more complex models (Cairns et al., 2009).

1.4.2 Projections of Mortality Rates

The principal objective of building extrapolative stochastic mortality models is to produce forecasts of the future rates along with the associated uncertainty levels. The generation of sufficiently many future paths constitute the projected distributions of mortality rates, which can then be summarised in terms of means and required percentiles. The dynamics of the models presented in this Chapter are driven by their time related parameters. The projections of those parameters are used to construct the paths of the future rates. Each family of models produces either central death or mortality rates. The outcomes might be compared consistently through relationship (1.1.7), which relates those indices. In this Section, the models are projected for a 50 year ahead horizon and the two families, namely the LC and CBD, are examined separately. For the LC family, the future distributions of $m(x,t)$ are displayed, while the CBD family of models is summarised in terms of $q(x,t)$.

The most important step before projecting the mortality models is the choice of the underlying stochastic model that govern their dynamics. Therefore, following the estimation of the latent states of the models, the second step of the traditional approach in stochastic mortality modelling is the selection of the appropriate models so that the projections are carried out. In principle and given the implied discrete time framework, the model selection should be based on formal Box-Jenkins methodology for the appropriate choice to be done (Box et al., 2013). The procedure consists of three main steps. Firstly, comes the model identification; this includes plot of the process along with plots of the estimated autocorrelation functions (ACF) and partial ACF's (PACF). Often, statistical tests are employed at this stage to examine underlying hypotheses. Once the appropriate model is identified, comes the step of estimating the relevant coefficients. The distributional assumptions about the model can then be used to carry out, for example, maximum likelihood estimation. Finally, comes the phase of diagnosing the chosen model. Commonly this is done by plotting residuals, ACF's and relevant QQ-plots. Again, various statistical tests may be used to verify the appropriateness of the chosen model. In the following paragraphs, we briefly comment on the choices of the relative models for the underlying stochastic factors.

Period Effects

As already commented, the original, and also more popular, choice for the stochastic period component of the LC model is random walk with drift. The model is strongly supported by the by the almost linear decrease of the $\kappa_t^{(2)}$'s with time for the majority of the examined cases. The volatility of the random walk, σ_κ, controls both the uncertainty in annual improvements and the size of accumulated deviations. This

particular model has been found to be appropriate for the majority of cases, and the process has also been maintained under several extensions and alternative forms of the original model (Renshaw and Haberman, 2003b, Pedroza, 2006). Even when a more complex model was found more relevant, it has been observed that the more sophisticated choice results in only slight differences (Lee and Miller, 2001). On the other hand, the random walk has been detected to change drift from time to time. This is the main motivation behind Booth et al. (2002)'s refinement of the estimation period for the model. However, such a method might lead to systematic underestimation of the true level of $\kappa_t^{(2)}$ (Cairns et al., 2008). Structurally different alternatives have been suggested by De Jong and Tickle (2006) who model the drift as a stochastic factor itself but under the state-space estimation framework, and by Czado et al. (2005) who use a mean-reverting process around a linear trend rather than a constant. In the projections conducted here, the period effects of the LC family of models are modelled by the originally suggested model so that:

$$\kappa_{t+1}^{(2)} = \kappa_t^{(2)} + \delta + \zeta_{t+1}, \tag{1.4.2}$$

where $\zeta_t \sim N(0, \sigma_\zeta^2)$, for all t and σ_ζ^2 is the variance of the disturbances.

Relative to the CBD family, following the original approach and subsequent analysis (Cairns et al., 2006a, 2011a), a multivariate random walk with drift is used to drive the dynamics of the models. Equivalently to before we have:

$$\boldsymbol{\kappa}_{t+1} = \boldsymbol{\kappa}_t + \boldsymbol{\delta} + \boldsymbol{\zeta}_{t+1},$$

where $\boldsymbol{\zeta}_t \sim N(0, V_\zeta)\ \forall t$, $\boldsymbol{\delta}$ is a constant drift vector and V_ζ is the covariance matrix, so that the disturbances of the random walk might be contemporaneously dependent but independent through time. Although, the model is questionable due to the possibility of producing biologically unreasonable forecasts, it has been found that in practical terms, and over a sensibly long term horizon, it behaves appropriately (Cairns et al., 2011a).

Cohort Effects

The estimates of the cohort effects derived from the *outer* cohorts of a given dataset are generally not regarded reliable. That is, the significant noise in the death counts along with the exposure to the relevant model risk, make these estimates less reliable as they are estimated from only a few observations (cells) of the data-set. Therefore, following Cairns et al. (2009, 2011a), cells with less than five observations have been excluded, so that the four more recent (1946-49) and the four earliest (1871-74) estimates produced by the model fitting are not taken into account.

Relative to the dynamics of the cohort parameters, the main widely accepted and used assumption is that they act independently of the corresponding period factors (Cairns et al., 2011a, Haberman and Renshaw, 2011). By the plots given in previous Sections it is clear that random walks would not be appropriate. The selection of a suitable model should be based on considering a variety of ARIMA models, and then choosing the one that fulfils the desirable criteria (highest BIC, fewer parameters, consistency with historical trends etc).

Regarding the $\gamma_c^{(3)}$ effects of the APC model, a variety of ARIMA models was considered for each dataset. Based on the resulting BIC, the best performing structures arise for the ARIMA(1,1,2) and ARIMA(5,1,1) specifications in the case of the EW dataset. On the other hand, the dominating models for the cohort process in the case of the Canada population were the ARIMA(1,1,2), ARIMA(2,1,2). For each case the ARIMA(1,1,0) model, used for the APC by Dowd et al. (2010b), Cairns et al. (2011a), is also considered. In the case of the EW population, beyond omitting the estimates of the outer cells, the parameters for years of birth 1918, 1919 and 1920 are also excluded, as discussed previously.

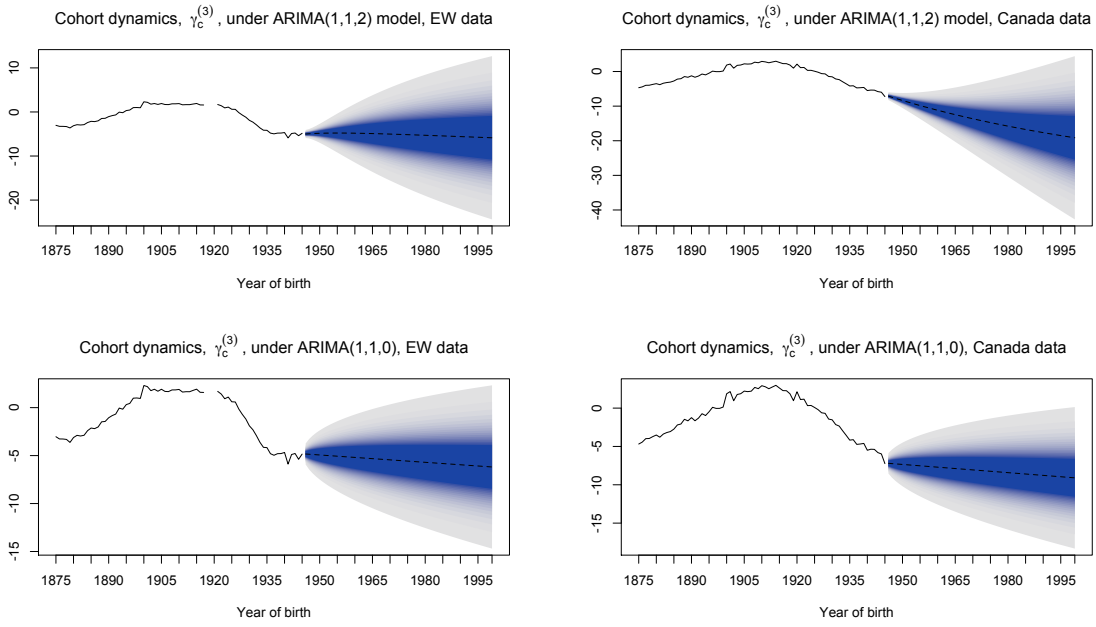

Figure 1.6: Estimated cohort process, $\gamma_c^{(3)}$, of the APC model for EW and Canada populations, along with fan charts of future dynamics under the assigned ARIMA specifications.

After analysing the residuals and future forecasts of the dominant models, the high order - ARIMA(5,1,1) and ARIMA(2,1,2) for the EW and Canada populations, respectively - are rejected as their complexity does not produce significantly distinct

forecasts to the simpler choices. The estimated $\gamma_c^{(3)}$ values and the associated fan charts of the projections are shown in Figure 1.6. The blue shaded regions indicate the 50% confidence level of the forecasts and each layer from thereon illustrates an additional 1% of confidence level up to the 99% interval. The ARIMA(1,1,2) model appears more flexible and seems able of generating future $\gamma_c^{(3)}$ factors of greater range compared to the ARIMA(1,1,0) model. Furthermore, the estimated residuals of the former model are more consistent with the underlying hypotheses, in contrast to the residuals of the ARIMA(1,1,0) model, which exhibit signs of heteroscedasticity. Additionally, the ACF's and PACF's of the residuals of the ARIMA(1,1,0) model indicate values that systematically exceed the allowed levels. The plots of the ACF's, PACF's and estimated standardised residuals of the ARIMA(1,1,2) model are plotted in Figure 1.7.

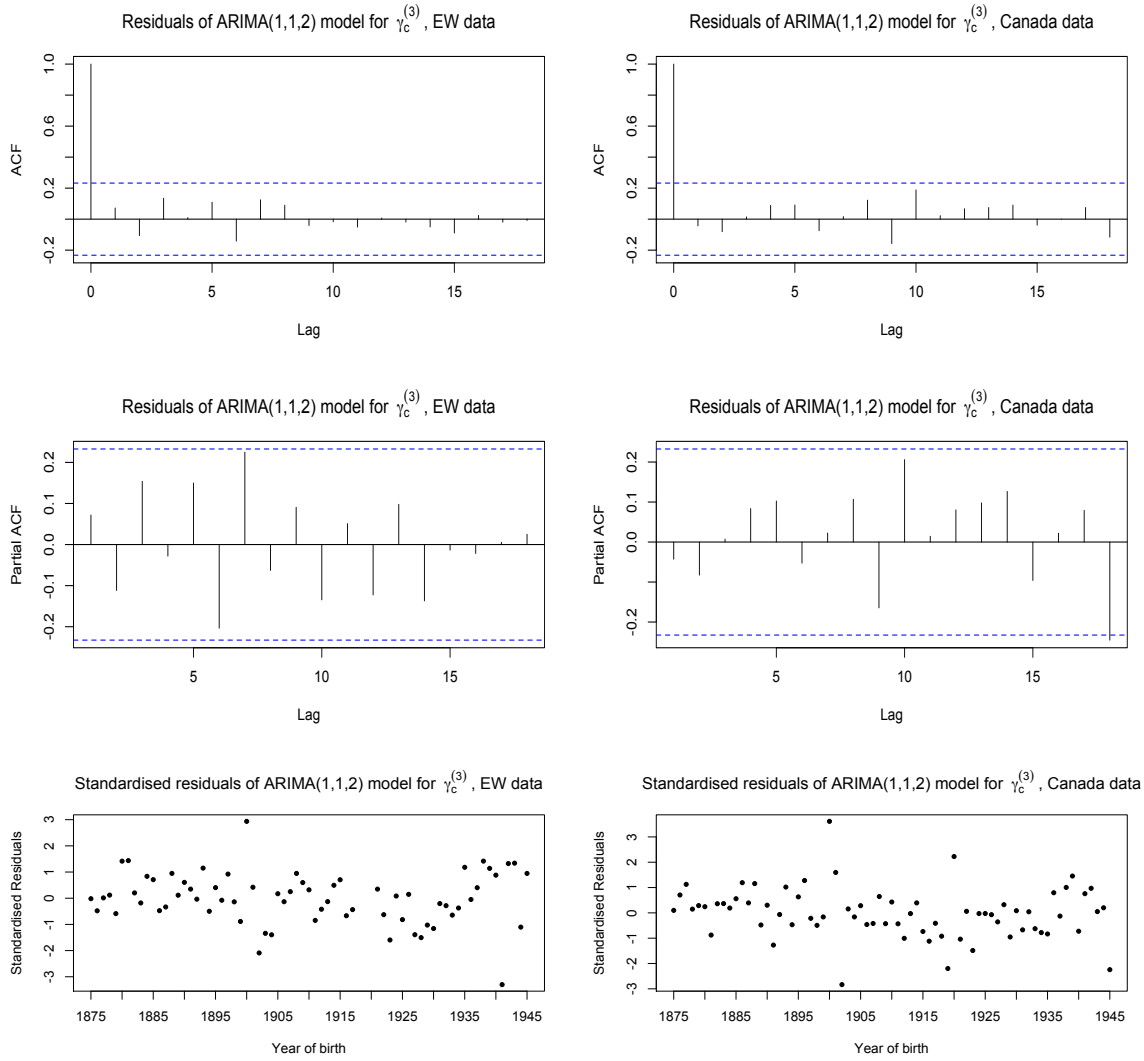

Figure 1.7: Residuals analysis of ARIMA(1,1,2) model for the cohort parameters, $\gamma_c^{(3)}$, of the APC model for EW and Canada populations, based on plots of estimated standardised values, ACF's and PACF's.

There are neither excessive amounts of autocorrelation nor of partial autocorrelation for both data-sets. The plotted standardised values appear within the required range and with no identifiable patterns. The APC model is projected using the ARIMA(1,1,2) and ARIMA(1,1,0) structures for the $\gamma_c^{(3)}$ effects. The two distinct models are labeled APC-MA and APC-MB, and will assist in recognising the degree of associated model risk between different specifications for the cohort effects.

For the cohort parameters of model M7, $\gamma_c^{(4)}$, the choice of the underlying stochastic model for the projection is more direct. The constraints applied to deduce the unique parameter estimates of the model imply that the fitted cohort parameters will fluctuate around 0 and will have no discernible linear trend or quadratic curvature. Thus, for model M7's $\gamma_c^{(4)}$ process an AR(1) model is used.

To summarise, the processes used for the cohort factors are given from the following equations. The ARIMA(1,1,2) process for $\gamma_c^{(3)}$ of the APC-MA model is:

$$\gamma_c^{(3)} = \delta_\gamma^{(3)} + (1 + \alpha_\gamma^{(3)})\gamma_{c-1}^{(3)} - \alpha_\gamma^{(3)}\gamma_{c-2}^{(3)} + \zeta_c^{(3)} + \zeta_{c-1}^{(3)} + \zeta_{c-2}^{(3)}.$$

The ARIMA(1,1,0) process for $\gamma_c^{(3)}$ of the APC-MB model is:

$$\gamma_c^{(3)} = \delta_\gamma^{(3)} + (1 + \alpha_\gamma^{(3)})\gamma_{c-1}^{(3)} - \alpha_\gamma^{(3)}\gamma_{c-2}^{(3)} + \zeta_c^{(3)},$$

and the AR(1) process for $\gamma_c^{(4)}$ of model M7 is:

$$\gamma_c^{(4)} = \delta_\gamma^{(4)} + \alpha_\gamma^{(4)}\gamma_{c-1}^{(4)} + \zeta_c^{(4)},$$

where the innovations $\zeta_i^{(j)} \sim N(0, \sigma^2_{i,\gamma^{(j)}})$, *i.i.d.* $\forall i$, and which are independent of the period disturbances for each model.

Forecasted Rates

Once the selection of the underlying processes for the stochastic factors of the mortality models has been established, the next step consists of producing the projections and comparing the outcomes. The forecasts serve as tools for inspecting the plausibility of the expected mortality term structure and the reasonableness of predicted levels of uncertainty compared to the realised history.

We examine the resulting paths of mortality rates for young, middle and old ages of the data-sets in hand. Thus, we plot the mortality statistics (the central death or mortality rate, depending on the LC or CBD family, respectively) for ages 65, 75, 85. Together with the mean expected forecasts, the 95% confidence bands are also shown, so that the relevant uncertainty is properly visualised. Based on these graphs, several observations might be made:

- Firstly, one of the drawbacks of the LC model, noted in Section 1.2.3, becomes apparent in Figures 1.8 and 1.10. That is, the confidence bands of the LC projections tend to narrow as the age of the examined group increases. That feature of the model can be considered as its major drawback, since it is in contrast both to one's common intuition and to previous realisations of mortality trajectories. To that respect, the CBD model is structurally different and yields distributions of future mortality rates which are increasingly volatile with age, as Figures 1.9 and 1.11 illustrate.

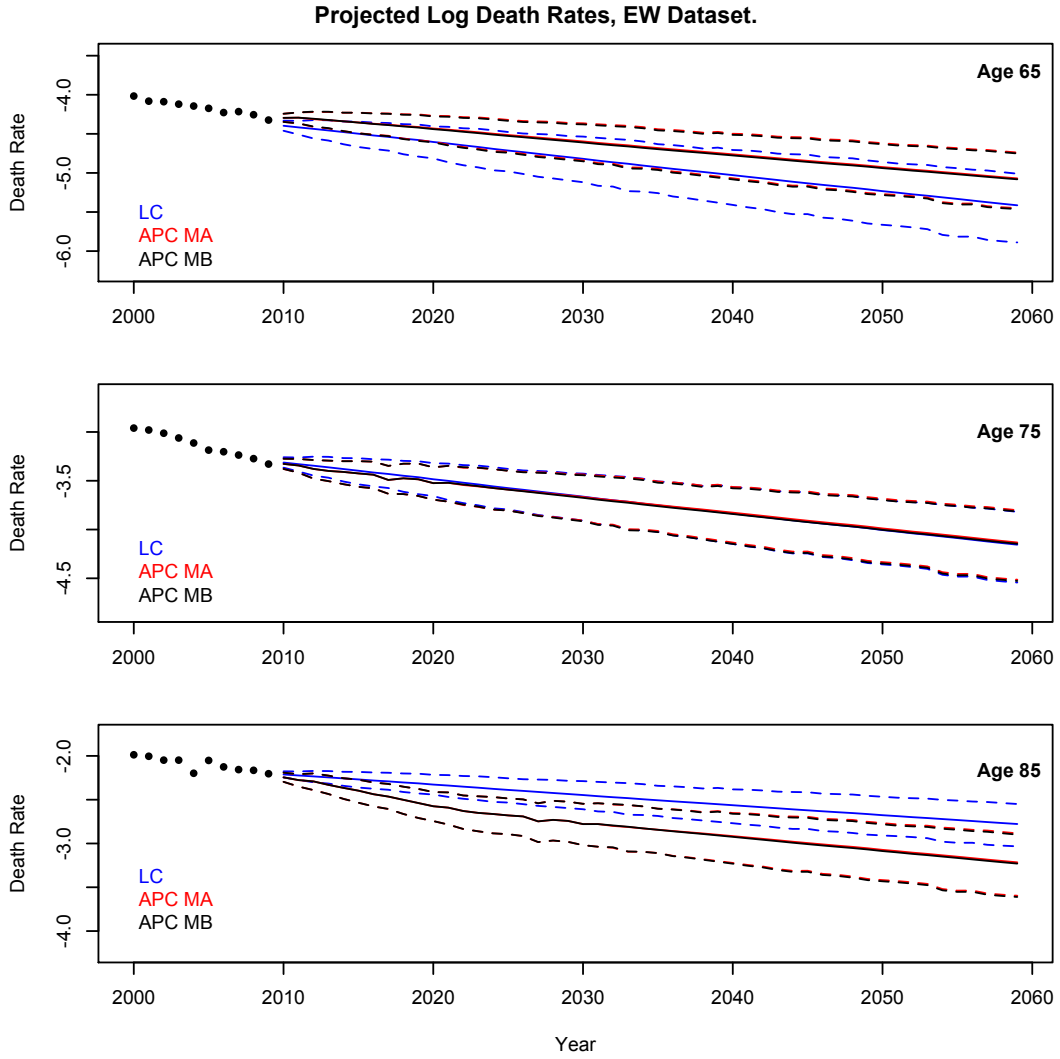

Figure 1.8: Projected central death rates on the log-scale for ages 65, 75 and 85 over a 50 years ahead horizon for the LC and APC models, EW data.

- For each family of models the projection graphs have been plotted in the identical range for the two populations. The comparison of the outcomes for fixed model and age group, but across populations, indicates that in the case of EW the forecasts are associated with significantly greater variability. This is

an indication of the distinctive behaviour of mortality in different data-sets. More specifically, the estimates of the period factors are generally smoother for the Canada data, hence the estimated uncertainty of the underlying stochastic model is lower than that of the EW population, and the produced confidence bands result in being narrower.

- A further observation that indicates the importance of different data-sets regards the way the APC model behaves under the APC-MA and APC-MB forms. Notably, the distinct specification for $\gamma_c^{(3)}$ between the two alternatives leads to almost indistinguishable differences in the case of the EW data. On the other hand, the different process for the cohort factors yields significant differences in the projections of the Canada population.

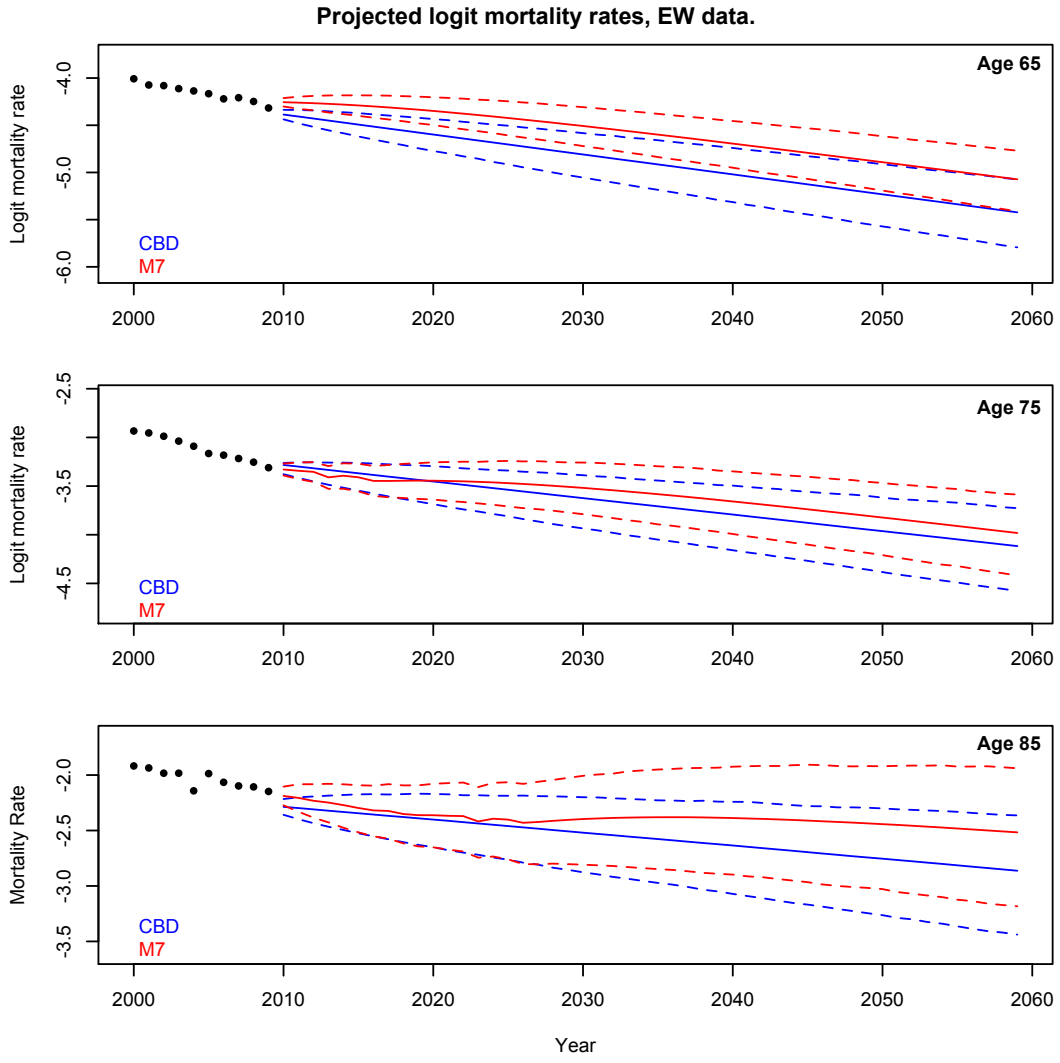

Figure 1.9: Projected mortality rates on the logit-scale for ages 65, 75 and 85 over a 50 years ahead horizon for the CBD and M7 models, EW data.

As soon as simulated cohort factors enter the projection, the forecasts develop notably different trends. This implies that different populations carry different degrees of model risk, so that the danger of misspecifying the cohort process when projecting the EW population is negligible compared to that arising from the Canadian data.

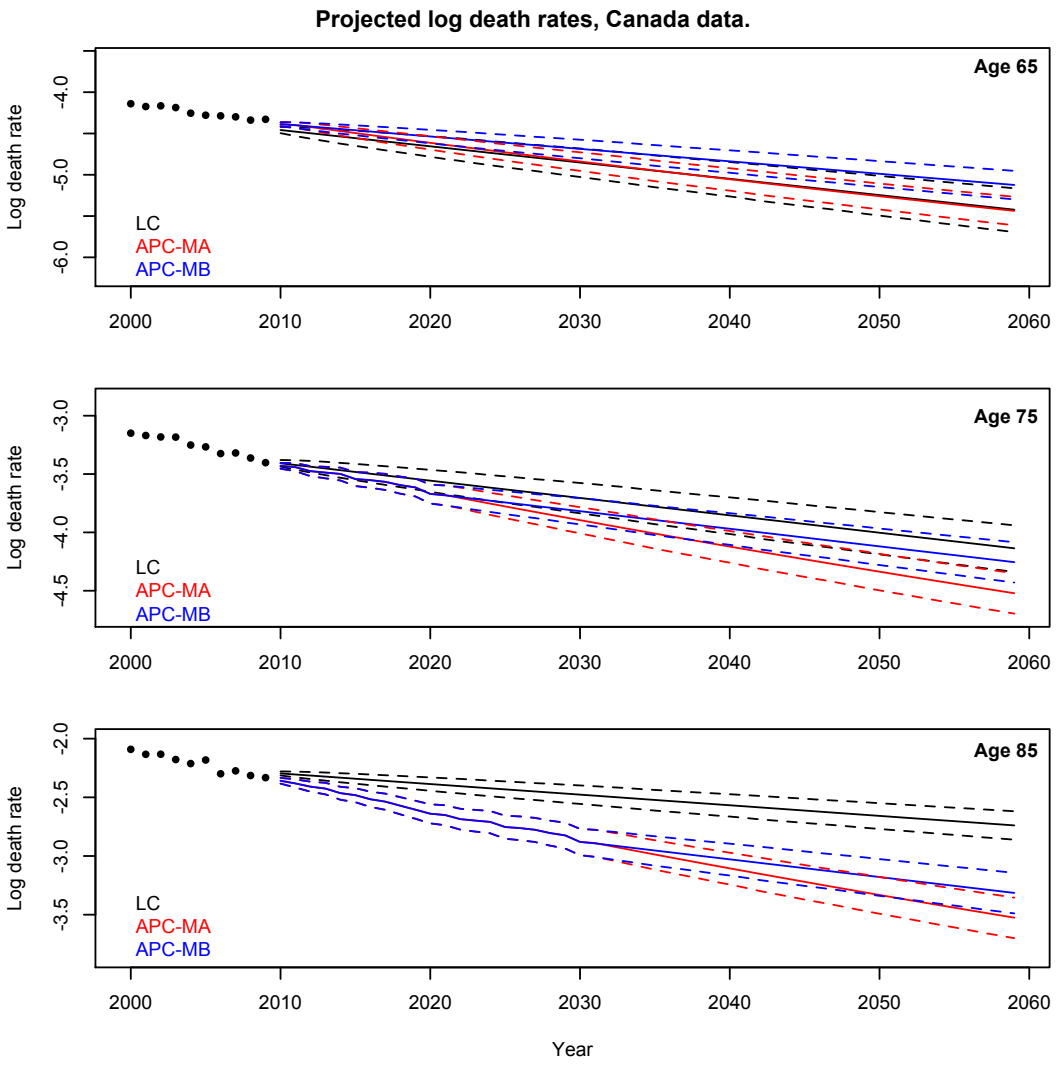

Figure 1.10: Projected central death rates on the log-scale for ages 65, 75 and 85 over a 50 years ahead horizon for the LC and APC models, Canada data.

- The models that have been enhanced with additional cohort (and in the case of M7, also period) factors yield wider confidence bands of future rates. This is expected and follows from the fact that the additional processes increase the uncertainty in the projections and thus the confidence regions around the central forecasts tend to get broadened. Models that include cohort effects develop jagged central forecasts for so long as estimated parameters are used to produce the forecasts. This is in accordance to the observed history of mortality

rates, but as soon as the estimates cease and projected parameter values come to play, the produced rates are smooth alike those of the non-cohort models. In turn, this questions the appropriateness of the stochastic models employed for the cohort parameters, since their projected values lead to distinct behaviour compared to their estimates.

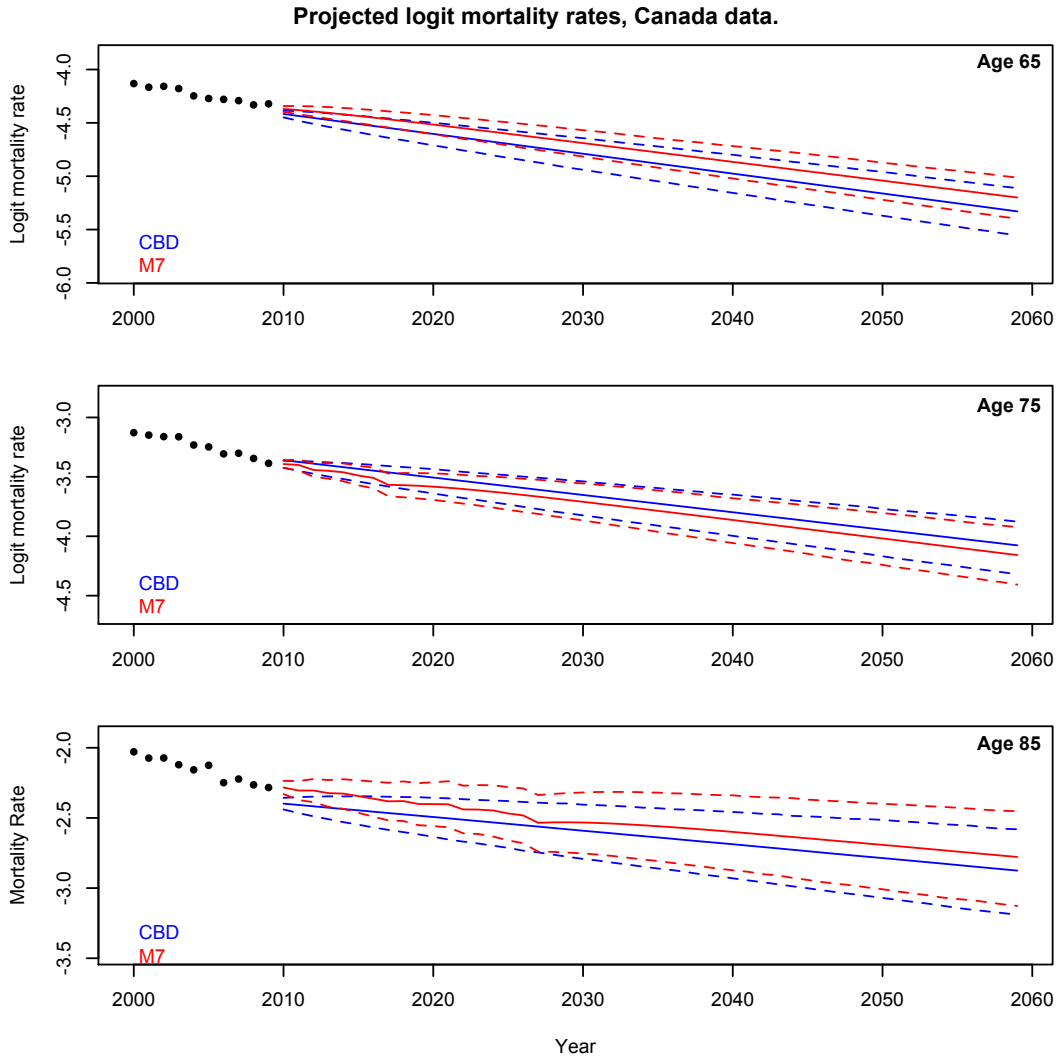

Figure 1.11: Projected mortality rates on the logit-scale for ages 65, 75 and 85 over a 50 years ahead horizon for the CBD and M7 models, Canada data.

The projections just examined provide useful points about how mortality models perform in terms of replicating the process based on historical data. The single factor of the LC model produces inconsistent results and the two- factor CBD structure appears more appropriate. Cohort enhanced models feed into the projections observed mortality patterns, but smooth out after the forecasts start incorporating projected cohort effects. Finally, the level of model risk for the cohort specification varies between the models and populations examined herein. The different quantitative

results between models APC-MA and APC-MB are only evident for the Canadian population and will be discussed at the end of the following Section. For further purposes of the Thesis, model APC-MA is taken as the cohort enhanced representative of the LC family, and henceforth the plain term APC model shall refer to the APC-MA version.

1.4.3 Further Projections and Numerical Examples

Based on the previously simulated mortality tables, the mortality metrics of Section 1.1.5 are calculated to examine further the forecasting properties of the implemented mortality models.

The projected survivor index, $S(65,t)$, for the next 25 years for all models and both data-sets are plotted in Figure 1.12. The plots compare the outcomes between models of the same family. Along the solid lines which indicate the mean forecasts, the shaded regions correspond to the 95% confidence bands. In the case of the LC family, the index deviates between the two models as the forecasting horizon widens. This is more intense and evident in the Canadian population where beyond 20 years ahead the forecasts do not overlap at all.

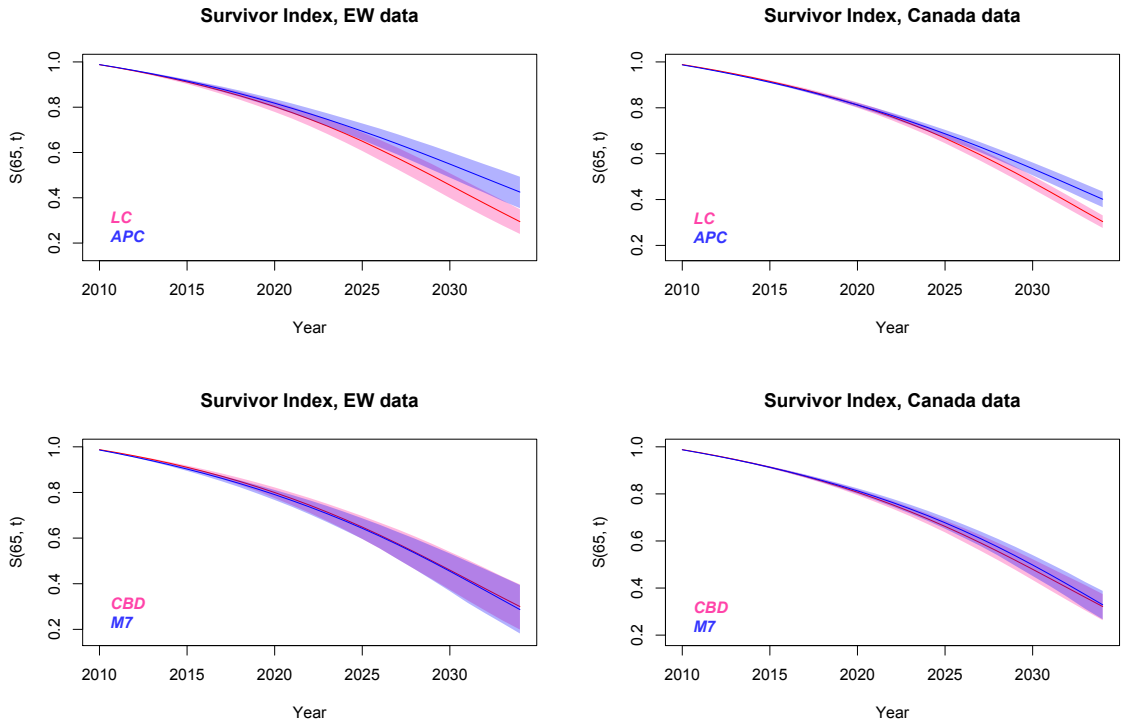

Figure 1.12: Projections of survivor indices, $\mathcal{S}(65,t)$, and associated confidence bands over the next 25 years for the LC, APC, CBD and M7 mortality models, EW and Canada data.

The main reason for the lack of overlap in the Canada dataset is the narrowness of the confidence bands, which follows from the respective plots and comments of the previous Section. Additionally, the APC model systematically predicts a higher proportion of the cohort staying alive by the end of the projection horizon for both data-sets. However, the mean forecast for $S(65, t)$ of the LC model is greater during the first decade of the simulation for the Canadian population. On the other hand, the results between the models of the CBD family are very comparable. The mean forecasts of the CBD model are greater than those of model M7 in the case of the EW data, but the converse is true for the case of the Canada population. The upper confidence levels of the CBD model are located slightly higher than those of model M7 for the EW data, and again the opposite holds for the Canada data-set. However, the general picture is that the projections under the CBD and M7 are very close amongst them for both populations, and that feature of the CBD family is in contrast to the LC models.

Based on the projection of the survivor index, the distributions of cohort life-expectancies and of annual term annuities payable in arrears for a maximum of 25 years are obtained. The life expectancies are calculated by assuming a limiting age of 100, and the required missing rates for the construction of the index are obtained via the projection method developed in Section 1.1.5. The results are presented in Table 1.3. The observed deviation of the mean survivor index projection between the members of the LC family leads to significantly different mean estimates of life-expectancies and annuity values for both data-sets. Additionally, the increased uncertainty levels of the EW population yield distributions with a higher spread for both metrics. The identical result holds for the uncertainty levels of the CBD family projections, with those of the Canada data exhibiting almost half the standard deviation for the examined mortality metrics of the EW population. Lastly, the closeness of the mean estimates for the survivor index of the two models of the CBD family produce mean estimates of the particular mortality metrics that are very similar.

The projected mortality table is also used to estimate the period life expectancies 25 and 50 years ahead, thus in years 2034 and 2059, as developed in Section 1.1.5. Table 1.4 includes the respective results. The APC model has been projected under both versions presented in Section 1.4.2 to assess the degree of model risk. In the case of the EW data, there are very small differences between the APC-MA and APC-MB models for both future time instances. However, the difference in the trend of the projections between the two versions of the APC model for the Canada population produce notably different life-expectancy results. Moreover, the detected differences appear to strengthen as the projection horizon extends. Notably, in the very far future of 50 years ahead, the APC-MA yields life expectancy distributions with almost

17% greater spread. The difference in the estimated standard deviations between the APC models to that of the LC model is much greater than the corresponding differences of the CBD family. Also, the APC models have greater life expectancy standard deviations than the M7 model, although the latter includes more stochastic factors. The uncertainty levels of the CBD family projections are close for the 25 years ahead projection under both data-sets. However, in the long term of 50 years ahead the impact of the additional stochastic components of model M7 determines the greater variability of simulated future lifetime distributions.

	EW data				Canada data			
Model	$\mathbb{E}(L)$	$Sd(L)$	$\mathbb{E}(P)$	$Sd(P)$	$\mathbb{E}(L)$	$Sd(L)$	$\mathbb{E}(P)$	$Sd(P)$
LC	19.37	0.5149	11.91	0.1843	19.60	0.2510	12.06	0.0935
APC	21.99	0.7499	12.40	0.1981	21.35	0.3547	12.30	0.1149
CBD	19.49	0.8984	11.91	0.2395	19.88	0.5240	12.04	0.1245
M7	19.13	0.9412	11.80	0.2423	19.96	0.5392	12.14	0.1296

Table 1.3: Means and standard deviations of cohort life-expectancies and of values of annual term annuities payable in arrears for a maximum of 25 years, P, for the LC, APC, CBD and M7 mortality models, EW and Canada data.

	EW data				Canada data			
	Year 2034		Year 2059		Year 2034		Year 2059	
Model	$\mathbb{E}(L)$	$Sd(L)$	$\mathbb{E}(L)$	$Sd(L)$	$\mathbb{E}(L)$	$Sd(L)$	$\mathbb{E}(L)$	$Sd(L)$
LC	21.81	1.1016	25.36	1.6383	21.76	0.5376	24.38	0.7318
APC-MA	24.58	1.9375	31.29	3.6729	26.21	0.9332	36.83	2.1227
APC-MB	24.66	1.9369	31.57	3.7173	25.63	0.9396	32.39	1.8095
CBD	22.04	1.8134	26.24	2.5015	22.35	1.0396	25.71	1.7139
M7	20.96	1.8829	23.30	3.1308	22.25	1.0623	24.92	1.9143

Table 1.4: Means and standard deviations of period life-expectancies in years 2034 and 2059 for the LC, APC, CBD and M7 mortality models, EW and Canada data.

1.5 Summary

This Chapter developed the construction of fundamental blocks and concepts underlying mortality modelling. After introducing some of the so-called early mortality laws and the intuition behind them, the most contemporary modelling assumptions were expounded . Based on these, benchmarks of the academic literature, namely the well known LC and CBD stochastic mortality models, along with cohort enhanced

extensions of them were exposed under a unified estimation framework. For each one of the illustrated models the accompanying maximum likelihood estimation algorithm was also given and the outcomes of the models for the EW and Canada data-sets were examined. In the final Section the stochastic mortality models were assessed in terms of goodness-of-fit and forecasting properties. Since the historical estimates of the stochastic factors appear generally smoother for the Canada population across all examined models, the confidence bands of the relevant forecasts are more narrow. Furthermore, the degree of model risk the two populations are subject to varies, with the projections of the Canadian population being more sensitive in the choice of the cohort effects process, $\gamma_c^{(3)}$, of the APC model. Additionally, the projected outcomes revealed that the majority of the implemented stochastic models behave quite similarly in terms of their direct forecasts and derivative quantities such as life-expectancies, with the exception of the APC model which produces the most erratic results. The results of this Chapter regarding the LC and CBD models will be revisited in Chapter 3, where the structures within their residuals will be identified and modelled appropriately to form the suggested extensions of the Thesis. The same forecasted results and mortality metrics will be also compared to those of the proposed models in Chapter 6.

Chapter 2

Bayesian Modelling and MCMC methods

This Chapter presents the fundamental elements of Bayesian modelling and introduces the concept of hierarchical models, which is related to the latent states mortality modelling set-up. We develop Markov chains Monte Carlo (MCMC) estimation algorithms, namely the Metropolis-Hastings and the Gibbs sampler along with several variants of them. The ideas behind hybrid algorithms are applied to the toy-example of the linear Gaussian State-Space model. Further, the matter of parameter uncertainty in the context of mortality modelling is discussed and the different approaches are reviewed. Finally, the CBD model is implemented under the Bayesian framework and the results are compared to those of the Poisson MLE solution of Chapter 1.

2.1 Introduction

The essence of Bayesian models lies in the fact that their parameters are treated as random quantities themselves, in contrast to the frequentist approach, in which they are supposed to have unknown fixed values. Supported by powerful computational tools, the Bayesian approach allows for a comprehensive treatment of the uncertainty underlying a statistical model.

2.1.1 Bayesian Analysis

Suppose the collected data $\boldsymbol{y}$ are intended to be modelled under some parametric form, dependent on the parameter vector $\boldsymbol{\theta}$. As with the frequentist statistical approach, the initially uncertain data $\boldsymbol{y}$ are described through a probability density, in the continuous case and without loss of generality, $f(\boldsymbol{y}|\boldsymbol{\theta})$. The observations of the vector $\boldsymbol{y}$ are typically assumed *i.i.d.* conditional on knowledge of $\boldsymbol{\theta}$, which determines the precise characteristics of the underlying distribution governing the model. The Bayesian paradigm formally incorporates knowledge or beliefs about the value of the parameter vector $\boldsymbol{\theta}$, through a density $p(\boldsymbol{\theta})$. The two ingredients described above are enough to consider the joint distribution of the data and the parameters of the model:

$$f(\boldsymbol{y}, \boldsymbol{\theta}) = f(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}). \tag{2.1.1}$$

The sampling distribution, determined by the density $f(\boldsymbol{y}|\boldsymbol{\theta})$, coincides with the likelihood of the parameter vector $\boldsymbol{\theta}$ under the classical statistical approach and provides the chances of each value of $\boldsymbol{\theta}$ having observed data $\boldsymbol{y}$. The distribution determined by the density $p(\boldsymbol{\theta})$, commonly known as the *prior distribution* of the parameters, supplies the model with an idea about $\boldsymbol{\theta}$ before observing any of the $\boldsymbol{y}$ values. Under the assigned prior specifications, it is plausible to make inference about $\boldsymbol{\theta}$ based on its probability distribution after observing the values of $\boldsymbol{y}$, that is, by using the *posterior distribution* of the model.

According to Bayes' law, the posterior distribution of $\boldsymbol{\theta}$, conditional on $\boldsymbol{y}$, is obtained as:

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{f(\boldsymbol{y}, \boldsymbol{\theta})}{f(\boldsymbol{y})} = \frac{f(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{f(\boldsymbol{y})}.$$

The density in the denominator is obtained by integrating equation (2.1.1) over the parameter space, so that:

$$f(\boldsymbol{y}) = \int_{\boldsymbol{\theta}} f(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}. \tag{2.1.2}$$

$f(\boldsymbol{y})$ is the marginal likelihood of the data, independent of the parameters $\boldsymbol{\theta}$ and is there to ensure that the integral of $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ over the whole parameter space of $\boldsymbol{\theta}$ is equal to 1. The density $f(\boldsymbol{y})$ is just a constant in terms of $\boldsymbol{\theta}$ and to that end, Bayes'

theorem may be compactly rewritten as:

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}). \tag{2.1.3}$$

The Bayesian approach yields imminent results about the uncertainty underlying the parameters of a given model, through their posterior distribution, $\pi(\boldsymbol{\theta}|\boldsymbol{y})$. Beyond the greater insight about the statistical model provided by the methodology, it is also often possible to increase the modelling flexibility by using distinct prior assignment for the underlying parameters.

2.1.2 Prior specification

The form of the posterior density, $\pi(\boldsymbol{\theta}|\boldsymbol{y})$, depends on the probabilistic model describing the data and on the prior assumptions made for $\boldsymbol{\theta}$. A prior is called conjugate for some form of likelihood if their combination through equation (2.1.3) results in a posterior of the same form as the prior. If this is the case, inference for the parameters of interest is easily done, since their posterior distribution is tractable.

For example, if one is interested in count data, a natural modelling assumption is that they are independently Poisson distributed, conditional on their distinct means, $\boldsymbol{\theta}$. Supplying this model with a Gamma prior for each of the elements of $\boldsymbol{\theta}$ results in the following well known Bayesian model.

$$\begin{aligned} Y_i|\theta_i &\sim Poisson(\theta_i) \\ \theta_i &\sim Gamma(a, b), \end{aligned} \tag{2.1.4}$$

for $i = 1, \ldots, n$.
The importance of conjugacy lies in the mathematical tractability of the required posterior. For the model of equation (2.1.4) it follows that:

$$\begin{aligned} \pi(\theta_i|\boldsymbol{y}) &\propto f(y_i|\theta_i)p(\theta_i) \\ &\propto \theta_i^{y_i} e^{-\theta_i} \theta_i^{a-1} e^{-b\theta_i} \\ &= \theta_i^{y_i+a-1} e^{-(b+1)\theta_i}. \end{aligned}$$

Therefore, the posterior distribution of θ_i, for $i = 1, \ldots, n$, is $Gamma(a + x_i, b + 1)$. The latter result allows exact inference about the parameters of the Poisson model, given the data.

The prior distribution is the point of divergence between the Bayesian and frequentist approaches. It might contain either previous historical information or some expert's opinion and knowledge. However, it might also assume ignorance about the precise features of the distribution of $\boldsymbol{\theta}$, so that the data reveal the real nature of the underlying processes. Ignorance about the parameters of the model is usually assigned through $p(\boldsymbol{\theta})$ by allowing $\boldsymbol{\theta}$ taking values with equal probability over a relevant interval. This might be done for instance if each element of $\boldsymbol{\theta}$ is uniformly distributed over some interval (l, u). If the interval is allowed to range over the whole real line, the prior is known as *improper uninformative* since its density does not integrate to 1. Equivalently, one can consider a Normal distribution with excessive spread as means of relaxing the strength of the prior information about the mean. Traditionally, non-informative improper prior specifications have given rise to plenty of controversy amongst Bayesians. This is in part due to the fact that such a specifications might lead to improper posterior distributions, that is, distributions that do not integrate to unity over their support.

A commonly accepted vague prior specification is the so-called Jeffreys' prior. Under the particular modelling assumptions, the Information matrix of the parameter vector is:

$$I(\boldsymbol{\theta}) = -\mathbb{E}\left[\frac{\partial \log\left[f(\boldsymbol{y}|\boldsymbol{\theta})\right]}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{'}} | \boldsymbol{\theta}\right].$$

Jeffreys' prior is then defined as:

$$p(\boldsymbol{\theta}) \propto |I(\boldsymbol{\theta})|^{1/2}.$$

Jeffreys' prior leads to appropriate prior forms for location and scale parameters, and is often a convenient uninformative choice.

2.1.3 Hierarchical Modelling

The parameters, $\boldsymbol{\phi}$, of the probabilistic model assigned to $\boldsymbol{\theta}$ are known as hyper-parameters of the model. In the case of the Poisson model of the previous section, the hyper-parameters are $\boldsymbol{\phi} = (a, b)$, which determine the Gamma prior. The fully Bayesian approach would include an additional model describing those parameters, in contrast to considering them as pre-specified or to-be-estimated constants. That would build up a hierarchical structure, which in analogy with section 2.1.1, contains the joint prior distribution of all the parameters $\boldsymbol{\phi}, \boldsymbol{\theta}$ of the model, and results in a joint full conditional posterior density of the form:

$$\pi(\boldsymbol{\phi}, \boldsymbol{\theta}|\boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{\phi}) p(\boldsymbol{\theta}, \boldsymbol{\phi}) = f(\boldsymbol{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta}|\boldsymbol{\phi}) p(\boldsymbol{\phi}),$$

since the likelihood usually does not depend on the parameterisation imposed on $\boldsymbol{\theta}$. The treatment is more than meaningful since, by further parameterising the $\boldsymbol{\theta}$ set, often more flexible assumptions might be made, while the uncertainty related with them is also assessed.

In the context that Bayesian theory adapts into mortality modelling one needs to add one more layer to the model due to the latent states explaining the data. For example, the mortality rates $q(x,t)$ under the CBD model are driven by the unobservable vectors of period effects, $\boldsymbol{\kappa}_t$. Therefore, the above developed hierarchical structure might be exploited to introduce the latent states required. If the latent variables are noted $\boldsymbol{x}$, then we are interested in the distribution $\pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})$. That posterior is characterised by the observation equation, $f(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})$, the latent states dynamics, $f(\boldsymbol{x}|\boldsymbol{\theta})$, and the prior $p(\boldsymbol{\phi})$ according to:

$$\pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}) f(\boldsymbol{x}|\boldsymbol{\theta}) p(\boldsymbol{\phi}).$$

In the case of the CBD model, the observation equation is the Poisson or the Binomial assumption for deaths, the latent state dynamics follow from the stochastic model for the $\boldsymbol{\kappa}_t$'s, the random walk with drift in this case, and any prior specifications about the parameters of the latter model constitute the prior of the model.

Hierarchical structures such as those described above most probably cancel any conjugate relationships. Moreover, the high dimension of the state vector might lead to problems where $\pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})$ is infeasible to be approached analytically. Computational methods known as Markov chain Monte Carlo algorithms may be used to generate samples from the target density.

2.2 Markov chain Monte Carlo methods

The loss or absences of conjugacy within a Bayesian model often makes the required posterior analytically intractable, and thus, not trivial to simulate from. As well as with the conjugacy cancelation within a hierarchical structure, one might choose for non-conjugate prior specification in order to gain flexibility. In such a case the marginal likelihood needs to be evaluated, in order to estimate the characteristics of the posterior distribution. In order to calculate the involved integrals, such as that of equation (2.1.2), it is possible to adopt asymptotic numerical approximations, such as Normal quadratures, or even include higher order terms resulting in a Laplace form of approximation of the integral. In a high dimensional parameter space these approximations often turn out to be inadequate (Gamerman and Lopes, 2006).

Alternatively, the use of simulation oriented methods may be considered to sample from the target posterior. The motivation behind such methods is to obtain a large enough sample from the posterior under consideration, and then use these values to summarise the distribution in terms of its location, dispersion and other measures of interest. Such methods are usually based on rejection or importance sampling techniques to draw relevant samples. They employ a sufficiently large sample of random variates from a single distribution. They are non-iterative schemes that depend highly on a sufficient approximation of the target distribution, which serves as the candidate generator.

It is often the case though that such an approximation is not possible to be found for the simulations to be carried out. In such situations it is feasible to sample from an appropriate Markov chain that converges to the target posterior. The essence of the method concentrates in the construction of the aforementioned Markov chain which would approximate the required posterior distribution. Iterating the chain for long enough leads to drawn values that correspond to the distribution of interest, and therefore, inference might be made based on the obtained converged sample. The method is widely known under the term Markov chain Monte Carlo (MCMC), highlighting the Markov chain based sequential nature of the technique. The method was originated under the context of statistical physics by Metropolis et al. (1953) and was further generalised by Hastings (1970). Hence, the most general algorithm is known as the *Metropolis-Hastings* (MH) scheme. A special case of the Metropolis-Hastings algorithm, known as the *Gibbs* sampler, has been explored by Gelfand and Smith (1990) in standard statistical problems. Further, Smith and Roberts (1993) and Tierney (1994) set up the underlying theoretical framework, while also developing its usefulness in Bayesian implementations. Textbooks by Carlin and Louis (2000), Gelman et al. (2003) and Gamerman and Lopes (2006) are considered classics for theoretical and practical considerations of MCMC algorithms.

2.2.1 Markov Chain theory

The aim of MCMC is the construction of a process which under an appropriate simulation scheme converges to some target distribution of interest. A Markov chain is a special kind of process built out of sequences of random variables. Suppose such a sequence of $d-$dimensional discrete random variables is $\{Y^{(n)} : n \in \mathbb{N}\}$, with associated state space $\mathcal{S}$ and index set the natural numbers $\mathbb{N}$. Since $\mathbb{N}$ is countable a discrete time process is implied, while $\mathcal{S}$ for illustration purposes is also assumed to be discrete, since the majority of essential asymptotic results have been initially established in a discrete state space. These might be analogously extended to more general state spaces.

The Markov property states that:

$$P\left(Y^{(n+1)} = y | Y^{(n)} = x, Y^{(n-1)} = x_{n-1}, \ldots, Y^{(0)} = x_0\right) = P\left(Y^{(n+1)} = y | Y^{(n)} = x\right),$$

for $x_0, \ldots, x_{n-1}, x, y \subset \mathcal{S}$.

If $Y^{(n)}$ is Markovian, the future development of the process depends on its current state, but every visited state before that is irrelevant. The transition probability of the chain, $P\left(Y^{(n+1)} = y | Y^{(n)} = x\right)$ depends on x, y and n. A chain for which the above probability does not depend on n is called *time homogeneous*. Such a homogeneous chain would have one step ahead transition probability:

$$P(x, y) = P\left(Y^{(1)} = y | Y^{(0)} = x\right),$$

which for all $x \in \mathcal{S}$ is a conditional probability distribution over $\mathcal{S}$.

For the case considered here, $P(x, y)$ is a matrix of probabilities, called the *transition matrix*. In the case of continuous chains it is replaced by a conditional density which defines the *transition kernel* of the chain.

The dynamics and limiting behaviour of Markov chains is of special interest. In particular, aggregating many steps of the chain we are interested in what is the m-step ahead transition probability:

$$P^m(x, y) = P\left(Y^{(m)} = y | Y^{(0)} = x\right).$$

Each transition often has also the meaning of iteration, so we are interested in what is the behaviour of the chain after iterating it for long enough. The fundamental result is that the chain will eventually forget its starting state, thus, $P^m(x, y)$ converges to a limiting distribution, with probability function say $\pi(\cdot)$, independent of the initial state x, as $m \to \infty$. Such a probability function $\pi(\cdot)$ expresses the *stationary* distribution of the chain and in the discrete case considered here should satisfy:

$$\sum_{x \in \mathcal{S}} \pi(x) P(x, y) = \pi(y), \quad \forall y \in \mathcal{S}. \tag{2.2.1}$$

For the convergence of a Markov chain to some *stationary* distribution with probability function $\pi(\cdot)$, three conditions must hold. The mathematical definitions of the following concepts are given in Appendix A.

Firstly, the chain has to be *irreducible*. Irreducibility implies that all the possible states are in communication within a finite number of transitions. Secondly, there must be no partition of the state space such that the chain can be constrained to cycle in these subsets from iteration to iteration. That is, the chain must exhibit

aperiodicity. Lastly, there must exist the relevant stationary distribution function $\pi(\cdot)$, reassured if the chain is *positive recurrent*. A state $x \in \mathcal{S}$ is *recurrent* if starting from state x the chain will return to x with probability 1. The state x is *positive recurrent* if the expected time of first return to x is finite. A Markov chain is called positive recurrent is all its states are *positive recurrent*. It can be shown that if the chain $Y^{(n)}$ is positive recurrent, there is a unique distribution function $\pi(\cdot)$ satisfying (2.2.1). Further, an aperiodic, positive recurrent chain is called *ergodic* and it holds that (Tierney, 1994):

$$\lim_{n\to\infty} P^{(n)}(x, y) = \pi(y), \quad \forall x, y \in \mathcal{S},$$

so that, iterating the chain for long enough leads to convergence to the stationary distribution.

2.2.2 Markov Chain Monte Carlo

Assuming that we know the target distribution $\pi(\cdot)$, given that the chain is aperiodic and irreducible, the problem is limited to appropriately specifying the probabilities $P(x, y)$ so that (2.2.1) is satisfied. To do that, the concept of reversible chains is facilitated. A chain is called *time reversible* if:

$$\pi(x)P(x, y) = \pi(y)P(y, x), \quad \forall x, y \in \mathcal{S}. \tag{2.2.2}$$

The point of the definition is that in equilibrium, the rate of transition from x to y is the same as the rate from y to x. Due to the balance in transition between the two states and since this holds for all states, condition (2.2.2) is also referred to as the *detailed balance* equations.

If a matrix $P(x, y)$ can be found so that equation (2.2.2) is satisfied, then summing (2.2.2) over y on both sides leads to equation (2.2.1), so that $P(x, y)$ is the required kernel. Therefore, the problem of constructing Markov chains with a given stationary distribution $\pi(\cdot)$ reduces to specifying $P(x, y)$ so that condition (2.2.2) holds.

First, the Metropolis-Hastings method for specifying appropriate kernels is developed. The convergence properties are proved by showing that the algorithm satisfies the detailed balance equations. The results underpinning MH are further applicable for the determination of the Gibbs sampler. These two techniques are presented next.

The Metropolis-Hastings algorithm

Focusing on the construction of the transition probabilities $P(x, y)$ so that the condition (2.2.2) holds, the irreducible and aperiodic chain $Y^{(n)}$ is updated as follows.

Given that the chain is in state x, generate a candidate value y from some proposal distribution $q(x, y)$ which is accepted as the next state of the chain with some probability $\alpha(x, y)$. The probability that the chain moves from x to y is:

$$P(x, y) = q(x, y)\alpha(x, y), \quad x \neq y.$$

Equivalently, the probability that the chain returns to x is:

$$P(x, x) = 1 - \sum_{y \neq x} q(x, y)\alpha(x, y).$$

Therefore, the transition probabilities of the Markov chain are given as:

$$P(x, y) = q(x, y)\alpha(x, y) + I(x = y)\left\{1 - \sum_{z \neq x} q(x, z)\alpha(x, z)\right\}, \quad \forall x, y \in \mathcal{S}, \quad (2.2.3)$$

where I is the indicator function, which takes value 1 if $x = y$ and 0 otherwise.

The requirement is for condition (2.2.2) to hold when the transition probabilities are given by equation (2.2.3). Condition (2.2.2) holds trivially when $x = y$, while for $x \neq y$, equation (2.2.3) becomes:

$$P(x, y) = q(x, y)\alpha(x, y),$$

and substituting this in to equation (2.2.2) yields:

$$\pi(x)q(x, y)\alpha(x, y) = \pi(y)q(y, x)\alpha(y, x), \quad \forall x, y \text{ s.t. } x \neq y. \quad (2.2.4)$$

Defining the acceptance probabilities $\alpha(x, y)$ to be:

$$\alpha(x, y) = min\left\{1, \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)}\right\},$$

equation (2.2.4) is satisfied, since for $\alpha(x, y) = \dfrac{\pi(y)q(y, x)}{\pi(x)q(x, y)}$ we get $\alpha(y, x) = 1$ and vice versa.

The Metropolis-Hastings algorithm, as presented above, produces a Markov chain that has stationary distribution the target distribution $\pi(\cdot)$. Its transition kernel is given by equation (2.2.3), where $q(x, y)$ is referred to as the *proposal distribution* and $\alpha(x, y)$ is the associated *acceptance rate* for candidates generated by the proposal $q(x, y)$. Despite being expounded for the case in which $Y^{(n)}$ is a discrete random variable, the method holds for continuous or mixed state spaces as well. If $\pi(\cdot)$ is multivariate, the scheme may be applied jointly, appropriately blocked or for each

individual component.

The appropriate choice of the proposal is a crucial step for the efficient performance of the algorithm. In theory it can be any distribution from which we can simulate, but appropriate choices will lead to faster convergence. *Symmetric* chains arise when $q(x, y)$ depends on (x, y) only through $|x - y|$. Then, $q(x, y) = q(y, x)$ and the acceptance rate reduces to:

$$\alpha(x, y) = min\left\{1, \frac{\pi(y)}{\pi(x)}\right\},$$

so that it is independent of the proposal $q(\cdot)$.

Random walk chains are a special case of symmetric chains. A random walk is a Markov chain, with evolution given as:

$$Y^{(n)} = Y^{(n-1)} + e_n,$$

where e_n is a random variable distributed independently of the chain. The disturbances are independent, identically distributed with density symmetric around zero. Proposed values are then based on the previous value of the chain. It follows that all the comments for symmetric chains are still valid leading to substantial computational simplifications. In practical implementations the dispersion of e_n is of high importance. Large dispersion leads to low acceptance rates, whereas, small variance would result in the chain moving slowly. The configuration of the proposal variance depends highly on the form of optimisation desired, while the scheme is easier to use component wisely.

Independence chains occur in the case which the transition is formulated independently of the previous state, so that $q(x, y) = g(y)$. It is generally recommended that g should be as close as possible to the target $\pi(\cdot)$, or at least it allows sampling from all probable values of it, otherwise the resulting sample will be misleading.

A pseudo code for joint sampling from a d–dimensional $\pi(\cdot)$, using the MH algorithm follows.

- Set the initial state $Y^{(0)}$
- For $n = 1, 2, 3, \ldots N$
  - Sample a proposed state y from $q(\cdot|Y^{(n-1)})$
  - Calculate the acceptance probability $\alpha = \dfrac{\pi(y)q(Y^{(n-1)}|y)}{\pi(Y^{(n-1)})q(y|Y^{(n-1)})}$
  - Generate a sample, u, from the random variable $U \sim U(0, 1)$
    - If $\alpha \geq u$ set $Y^{(n)} = y$, otherwise $Y^{(n)} = Y^{(n-1)}$.

- End for.

The convergent sample of the above algorithm is taken as the required sample from the target posterior distribution. As already mentioned, the acceptance rate is an important component of the scheme, and a useful tool for trial and error adjustments both with the form and the specific characteristics of the proposal distribution. As a rule of thumb the optimal convergence properties for specific problems range from 25% to 40%. However, the dimensionality of the problem is an important component, and in complicated state spaces those suggestions may not be very realistic.

The Gibbs sampler

The Gibbs sampler is a special case of the MH algorithm. Consider sampling from the d–dimensional $\pi(\cdot)$, and denote by $\pi_i(\cdot)$ the distribution of the $i-th$ component conditional on the other $d-1$. This is called the full conditional distribution of the $i-th$ component, and if it is recognisable from the setting of the problem, it is used as a proposal leading to a scheme with acceptance rate equal to 1.

A pseudocode of the Gibbs algorithm for element wise sampling from $\pi(\cdot)$ via the chain Y is as follows:

- Initialise at $Y^{(0)} = (Y_1^{(0)}, \ldots, Y_d^{(0)})$
- For $n = 1, 2, 3, \ldots N$ draw sequentially samples:
  - $Y_1^{(n)} \sim \pi_1(Y_1 | Y_2^{(n-1)}, \ldots, Y_d^{(n-1)})$
  - $Y_2^{(n)} \sim \pi_2(Y_2 | Y_1^{(n)}, Y_3^{(n-1)}, \ldots, Y_d^{(n-1)})$
  - $\vdots$
  - $Y_k^{(n)} \sim \pi_k(Y_k | Y_1^{(n)}, \ldots, Y_{k-1}^{(n)}, Y_{k+1}^{(n-1)}, \ldots, Y_d^{(n-1)})$
  - $\vdots$
  - $Y_d^{(n)} \sim \pi_d(Y_d | Y_1^{(n)}, \ldots, Y_{d-1}^{(n)})$.
- End for.

Therefore, the scheme reduces in sampling from the full conditionals provided by the model. When this is not possible, it might still be used with appropriate approximations to the full conditionals. Under mild regularity conditions the specification of all the full conditional distributions uniquely determines the full joint distribution and hence, all the marginals. This is a special case of the Hammersley-Clifford theorem (Besag, 1974, Johannes and Polson, 2009), and in terms of the latent state hierarchical structure of the previous section, its essence lies in the fact that knowledge of $f(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$ and $f(\boldsymbol{\theta}|\boldsymbol{x}, \boldsymbol{y})$ completely characterise the joint target density $\pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})$. Further, given that $Y^{(n)}$ is aperiodic and irreducible, it follows from the more general MH case that as $n \to \infty$, $Y^{(n)} \to \pi(\cdot)$ (Gamerman and Lopes, 2006).

2.2.3 Practical Considerations

Although the theory behind MCMC algorithms for Bayesian inference is straightforward, there are still issues one must be aware of to avoid relevant pitfalls or to improve the performance of the implemented solutions. Nevertheless, the convergence properties of such methods strongly rely on the nature of the problem under question and the chosen method. For example, the addition of latent variables in the state-space can dramatically improve the rate of convergence (Polson, 1996). On the other hand, improper uninformative prior distributions may lead to improper posteriors which would then result in a degenerate MCMC algorithm. Examples of when this might occur may be found in Hobert and Casella (1996).

Values from the target distribution $\pi(\cdot)$ are supposed to be obtained once the chain has converged to its steady state. However, the determination of how long it should be iterated is in no way pre-specified and depends on the problem at hand. A standard way that may improve significantly the convergence speed of an MCMC algorithm is by blocking correlated parameters and sampling them jointly, even though blocking will usually result in increased posterior correlation. If the model is intended to be projected the correlation between its parameters will feed into the projections and cause bias. Hence, blocking should be used wisely so that the relevant projections are not adversely influenced. From a statistical point of view, the convergence of an MCMC algorithm might be examined with a number of assessment criteria. Cowles and Carlin (1996) provide a comparative review of several such practical diagnostics under various different settings. They are based either on multiple runs of the same chain or on single chains, while some of them are only available for specific simulation schemes, *e.g.* Gibbs sampler. A lot of these tools are readily implemented in Bayesian output analysis packages, such as the CODA by Plummer et al. (2006) or the BOA by Smith (2007).

Such a diagnostic that will be used in the following Sections and Chapters was developed by Geweke (1992) and it is based on the spectral properties of the posterior sample. If the MCMC sample of some parameter is denoted as $\theta^{(j)}$ for $j = 1, \ldots, n$ and two subsequences of that sample, an initial of size $n_A = k_1 n$ and a final of size $n_B = k_2 n$, are chosen such that

$$\frac{n_A + n_B}{n} < 1,$$

and if the sequence $\theta^{(j)}$ is stationary, then

$$Z_n = \frac{\bar{\theta}_A + \bar{\theta}_B}{\sqrt{\frac{1}{n_A}\hat{S}_\theta^A(0) + \frac{1}{n_B}\hat{S}_\theta^B(0)}} \to N(0,1) \text{ as } n \to \infty,$$

where

$$\bar{\theta}_A = \frac{1}{n_A}\sum_{j=1}^{n_A}\theta^{(j)},\ \ \bar{\theta}_B = \frac{1}{n_B}\sum_{j=1}^{n_B}\theta^{(j)}$$

and $\hat{S}_\theta^A(0)$ and $\hat{S}_\theta^B(0)$ are the spectral estimates at 0, which are also consistent estimators for the true variance of θ (Brooks and Roberts, 1998). The values of k_1 and k_2 were originally prescribed to be such that $k_1 = 0.1$ and $k_2 = 0.5$, although any sensible adjustment to these figures is also acceptable. The diagnostic was originally intended to determine the amount of initial iterations that should be discarded. However, it can also be applied to an already seemingly converged sample as a means of verification.

One of the most critical components of an MCMC algorithm is the proposal distribution. In theory, any candidate generator would suffice for the convergence of the algorithm in the long run. However, and especially in multidimensional state-spaces, a proposal which is close to the target posterior may improve dramatically the convergence rate of the implemented chain. Adaptive algorithms are common tools employed for the optimisation of simpler MCMC schemes. Theory and practical considerations about adaptive algorithms may be found in Andrieu and Thoms (2008) and Atchade et al. (2009). Any algorithm which sequentially updates, or just adjusts its internal parameters according to some criterion, may be considered as adaptive. All the algorithms developed from Section 2.4.1 and onwards include some sort of adaptation. Another practical way to get improved estimates within a hybrid algorithm if an adequate proposal distribution is not available is to iterate the individual scheme more, and estimate the parameters and the acceptance rates by the respective averages. In all cases, the performance of an MCMC scheme highly depends on its construction and the subjective qualitative and quantitative choices adopted.

Once the MCMC sample is obtained one checks the trace-plots of the parameters of interest and typically, discards some amount of initial iterations. The discarded iterations are known as the *burn in* period of the chain. Further, it is desirable that consecutive iterations of the chain constitute an *i.i.d.* sample from the target posteriors. Under the time series perspective of the chains for individual parameters this might be assessed through their ACF's. If there is high autocorrelation in the examined samples, one might opt for *thinning* the chain, so that only every *k-th* iteration is held. Lastly, multi-parameter models should exhibit a low level of cross-correlation between their parameters. Although this is difficult in cases which the parameters are connected with the same variable, re-parameterisation techniques might be applied to reduce the effect. Once a satisfyingly convergent sample has been achieved, statistical and visual analysis is easily done, summarising measures of central tendency, dispersion and further quantities of the chain(s) of interest.

2.3 MCMC for State Space Models

As an application of the Bayesian paradigm, a bi-dimensional multi-parameter State Space Model is simulated and then estimated back. The exercise serves as an example of how the techniques developed in the present Chapter may be applied.

2.3.1 Model Specification

Assume there are observations in two dimensions, $\boldsymbol{y} = y(i,j)$, generated as follows:

$$\begin{aligned} y(i,j) &= x(i) + \sigma_y \zeta(i,j) \\ x(i) &= x(i-1) + \mu + \sigma_x \omega(i), \end{aligned}$$

where $\zeta(i,j), \omega(i)$ are *i.i.d.* Standard Normal variates for $i = 1, \ldots, n$ and $j = 1, \ldots, m$.

The above model is a special type of a linear Gaussian state-space model. The vector of unknowns, $\boldsymbol{x}$, describes the unobservable state of the system, and hence is called the state equation with *i.i.d.* $N(0, \sigma_x^2)$ innovations. The equation for $\boldsymbol{y}$ provides the link between the observed data and the states $\boldsymbol{x}$. It is known as the observation equation under measurement error with variance σ_y^2. The random walk $\boldsymbol{x}$ might be interpreted as the latent local trend driving the observations $\boldsymbol{y}$. The parameters to be estimated are the latent random walk, $\boldsymbol{x}$, along with its parameters, μ, σ_x, and the variance of the noise associated with the observations $\boldsymbol{y}$.

2.3.2 Bayesian Estimation

The ideas behind hierarchical modelling developed in section 2.1.3 may be applied to the estimation of the latent states $\boldsymbol{x}$. Essentially, filtering the states $\boldsymbol{x}$ out of the observations $\boldsymbol{y}$ provide all the information required to estimate the parameters μ and σ_x^2, which determine the dynamics of the system. Let the global parameter vector be $\boldsymbol{\omega} = \left(\boldsymbol{x}, \sigma_y^2, \mu, \sigma_x^2\right)$ and also let $\boldsymbol{\theta} = \{\mu, \sigma_x^2\}$. The full set of distributional assumptions of the hierarchical model are as follows:

$$\begin{aligned} y(i,j)|x(i), \sigma_y^2 &\sim N\left(x(i), \sigma_y^2\right) \\ x(i)|x(i-1), \mu, \sigma_x^2 &\sim N\left(x(i-1) + \mu, \sigma_x^2\right). \end{aligned}$$

In order to conduct the exercise, observations are simulated under the specified structure. The fixed parameter values are $\mu = 0.03, \sigma_x^2 = 4 \times 10^{-4}, \sigma_y^2 = 16 \times 10^{-4}$ and the random walk was initiated at $x(1) = 0$. Once the random walk was simulated, each of

the $x(i), i = 1, \ldots, n$ values is used to simulate the $y(i, j)$ points that serve as observations, for $i = 1, \ldots, n,\ j = 1, \ldots, m$. The simulated data are used to estimate the model by implementing a hybrid MCMC algorithm. The following prior assumptions are specified for the parameters of the model:

$$\sigma_y^2 \sim IG(a, b), \mu \sim N\left(\mu_0, \sigma_\mu^2\right), \sigma_x^2 \sim IG(c, d), \tag{2.3.1}$$

with $\mu_0 = 0, \sigma_\mu^2 = 1, a = b = c = d = 0.001$. The IG symbol indicates the Inverse Gamma distribution which is the well-known conjugate pair of the variance of the Normal distribution. Throughout the Thesis a random variable variate, X, from the IG distribution has density of the following form:

$$f(x) = \frac{b^a}{\Gamma(a)} x^{-(a+1)} \exp(-\frac{b}{x}),$$

where $\Gamma(\cdot)$ is the Gamma function. The common small value for the parameters of the IG distributions is a typical choice to indicate prior ignorance while resulting in analytically tractable full conditional posteriors. Under prior independence between all three model parameters μ, σ_x, σ_y, we are interested in the posterior distribution:

$$\pi(\omega|\boldsymbol{y}) \propto f\left(\boldsymbol{y}|\boldsymbol{x}, \sigma_y^2\right) \times f\left(\boldsymbol{x}|\boldsymbol{\theta}\right) \times p(\boldsymbol{\theta}, \sigma_y^2).$$

$f(\boldsymbol{y}|\boldsymbol{x}, \sigma_y^2)$ corresponds to the Normal likelihood of the data $\boldsymbol{y}$, $f\left(\boldsymbol{x}|\boldsymbol{\theta}\right)$ describes the dynamics of the latent states $\boldsymbol{x}$, so that:

$$f\left(\boldsymbol{x}|\boldsymbol{\theta}\right) = \prod_{i=2}^{n} f\left(x(i)|x(i-1), \boldsymbol{\theta}\right).$$

Finally, $p(\boldsymbol{\theta}, \sigma_y^2)$ is the joint prior for the parameters of the model.

The prior distribution of the mean is concentrated around the true value of the parameter but the exercise was also done with relatively higher values of prior variance with no significant effects.

The Normal distributions involved, equipped with the appropriate prior specifications yield a conjugate structure. Define $\boldsymbol{\omega}_{-i}$, the global parameter vector except for the parameter considered. Then the full conditional posteriors of $\mu, \sigma_x^2, \sigma_y^2$ are as follows:

$$f(\mu|\boldsymbol{\omega}_{-i}) \sim N\left(\frac{\sigma_\mu^2 \sum_{j=2}^{n}\left(x(j) - x(j-1)\right) + 2\mu_0\sigma_x^2}{n\sigma_\mu^2 + \sigma_x^2}, \frac{\sigma_\mu^2\sigma_x^2}{n\sigma_\mu^2 + \sigma_x^2}\right)$$

$$f(\sigma_x^2|\boldsymbol{\omega}_{-i}) \sim IG\left(a + \frac{n}{2}, b + \frac{1}{2}\sum_{j=2}^{n}\left(x(j) - x(j-1) - \mu\right)^2\right)$$

$$f(\sigma_y^2|\boldsymbol{\omega}_{-i}) \sim IG\left(c + \frac{nm}{2}, d + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}(y(i,j) - x(i))^2\right).$$

Similar Normal full conditional posteriors might be derived for the latent states of the random walk. Instead, random walk MH steps are implemented to sample from them, sequentially and conditionally on the latest update of each predecessor. The complete algorithm estimates the model very accurately and convergence is achieved rapidly. The Normal proposal of the Metropolis step of the chain is assigned variance 9×10^{-4}, leading to an acceptance rate of around 21% across all latent states, $\boldsymbol{x}$. The algorithm runs for 31000 iterations, of which 1000 are discarded as burn–in period. Thinning the sample so that every 10–th iteration is kept, results a posterior sample of size 3000 for the parameters of interest, which is enough to produce samples that exhibit no significant amount of autocorrelation.

2.3.3 Convergence and summary of the posterior distribution

The convergence of the algorithm is assessed by visual and statistical diagnostics, as discussed in Section 2.2.3. We examine the convergence and sample properties of the parameters of the model, μ, σ_x^2 and σ_y^2, along with the indicative states of the random walk, x_6, x_{16} and x_{26}. Firstly, the obtained samples are tested with Geweke's diagnostic. The means of the first 20% of the chains are compared against that of the final 40%. The resulting Z-scores and the corresponding p-values are shown in Table 2.1. Across all estimated parameters, there are only two cases in which the underlying hypothesis of the test could be rejected at the 5% confidence level and one of them is for the state x_{26} as shown in the table below.

	μ	σ_x^2	σ_y^2	x_6	x_{16}	x_{26}
Z-score	0.604	-1.324	1.109	.966	0.233	-1.961
p-value	0.546	0.186	0.268	0.334	0.816	0.049

Table 2.1: Z-scores and p-values of Geweke's diagnostic for parameters μ, σ_x^2 and σ_y^2 and for latent states x_6, x_{16} and x_{26} of the Gaussian State Space model.

The cumulative trace-plots of the means and of the associated 95% highest posterior density (HPD) bands are shown in Figure 2.1. The red dashed lines indicate the true values of the parameters and latent states. The convergence appears fairly satisfying for all examined cases. Particularly, the estimation accuracy of the algorithm is better for the latent random walk, $\boldsymbol{x}$, rather than for the parameters of the model. However, the parameters of the random walk, μ and σ_x^2, seem to be estimated better

than the variance parameter, σ_y^2, which appears noticeably shifted. Additionally, the trace-plot for σ_x^2 indicates a strongly skewed distribution, which follows since it is sampled from an IG distribution. On the other hand, the posterior density of σ_y^2 appears fairly symmetrical, but it is also the one with the greater deviation from the true value.

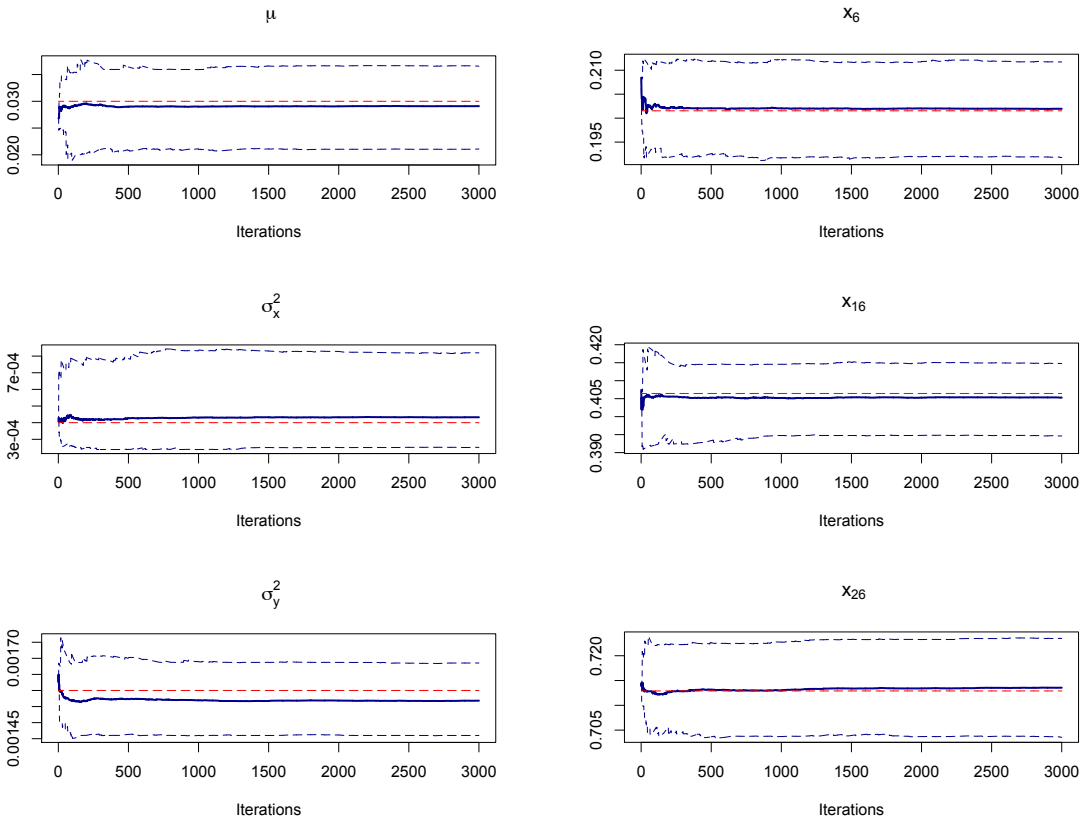

Figure 2.1: Cumulative trace-plots of means and of the associated 95% HPD intervals for parameters μ, σ_x^2 and σ_y^2 and for latent states x_6, x_{16} and x_{26} of the Gaussian State Space model.The red dashed lines indicate the original values of parameters and latent states.

Table 2.2 summarises the means, their standard errors and the standard deviations of the parameters of the model and of the considered estimated latent states. There are also tabulated the lower and upper bands of the 95% HPD intervals. For each parameter the interval is constructed from the empirical cumulative distribution function (ECDF) of the sample as the shortest interval for which the difference in the ECDF values of the endpoints is the nominal probability (Plummer et al., 2006). Table 2.2 might be visualised by the plots of the posterior densities in Figure 2.2, along with the credibility intervals in dashed lines and the original values of the parameters and latent states in red.

The results of the convergence assessment show that the Gaussian State Space model is captured efficiently by the implemented MCMC algorithm. However, estimates of the latent states appear more accurate relative to those of the parameters of the model, especially when considering the variance, σ_y^2. This might be an effect of the hierarchy of the model reflected in the estimation precision. The latent states $\boldsymbol{x}$ are higher in the hierarchy of the model and all of the full conditional posteriors of μ, σ_x^2 and σ_y^2 depend on the successive iterations of the $\boldsymbol{x}$ vector.

As mentioned in Section 2.2.3, the output of an MCMC algorithm should constitute an *i.i.d.* sample of the target posterior distribution. However, in some cases the marginal chains of parameters exhibit a significant amount of autocorrelation. Thinning the chain is the most effective way of dealing with this sort of problem. The autocorrelation functions of the examined estimated quantities of this section are shown in Figure 2.2.

	Mean	Mean s.e.	Std Deviation	L_{hpd}	U_{hpd}
μ	0.02946	7.091×10^{-5}	3.884×10^{-3}	0.02157	0.03681
σ_x^2	3.949×10^{-4}	2.575×10^{-6}	1.411×10^{-4}	2.2064×10^{-4}	7.34×10^{-4}
σ_y^2	15.682×10^{-4}	1.062×10^{-6}	5.819×10^{-5}	14.564×10^{-4}	16.795×10^{-4}
x_6	0.2025	1.055×10^{-4}	5.110×10^{-3}	0.1932	0.201
x_{16}	0.406	1.011×10^{-4}	5.197×10^{-3}	0.3966	0.4134
x_{26}	0.7133	9.669×10^{-5}	5.143×10^{-3}	0.7058	0.7227

Table 2.2: Posterior summary for parameters μ, σ_x^2 and σ_y^2 and for latent states x_6, x_{16} and x_{26} of the Gaussian State Space model.

The parameters μ, σ_x^2 and σ_y^2 appear to have no sign of statistically significant amount of autocorrelation at any lag. On the other hand, the estimated latent states indicate systematic significant autocorrelation at the first lag. A potential isolation to that would be to thin further the obtained sample or rerun the algorithm more times and thin it at a higher lag from scratch. However, since this is merely an illustrative example no further care is taken for that issue.

If the model were to be projected in the future, it would also be essential for the sample not to exhibit cross-correlation between the individual parameters of the model. Otherwise, the projections would carry forward the correlation between the parameters of the model. The cross-correlation between the parameters is measured by the correlation matrix of the MCMC series of μ, σ_x^2 and σ_y^2. The respective figures are $(\rho_{\mu,\sigma_x^2} = -0.0059, \rho_{\mu,\sigma_y^2} = 0.00276, \rho_{\sigma_x^2,\sigma_y^2} = -0.0135)$, which are negligible. The cross-correlation between the latent states, $\boldsymbol{x}$, is shown through the contours of the corresponding matrix in Figure 2.3. The identical observation as for the triple $(\mu, \sigma_x^2, \sigma_y^2)$

is true for those, and in summary, the posterior sample does not bring up any

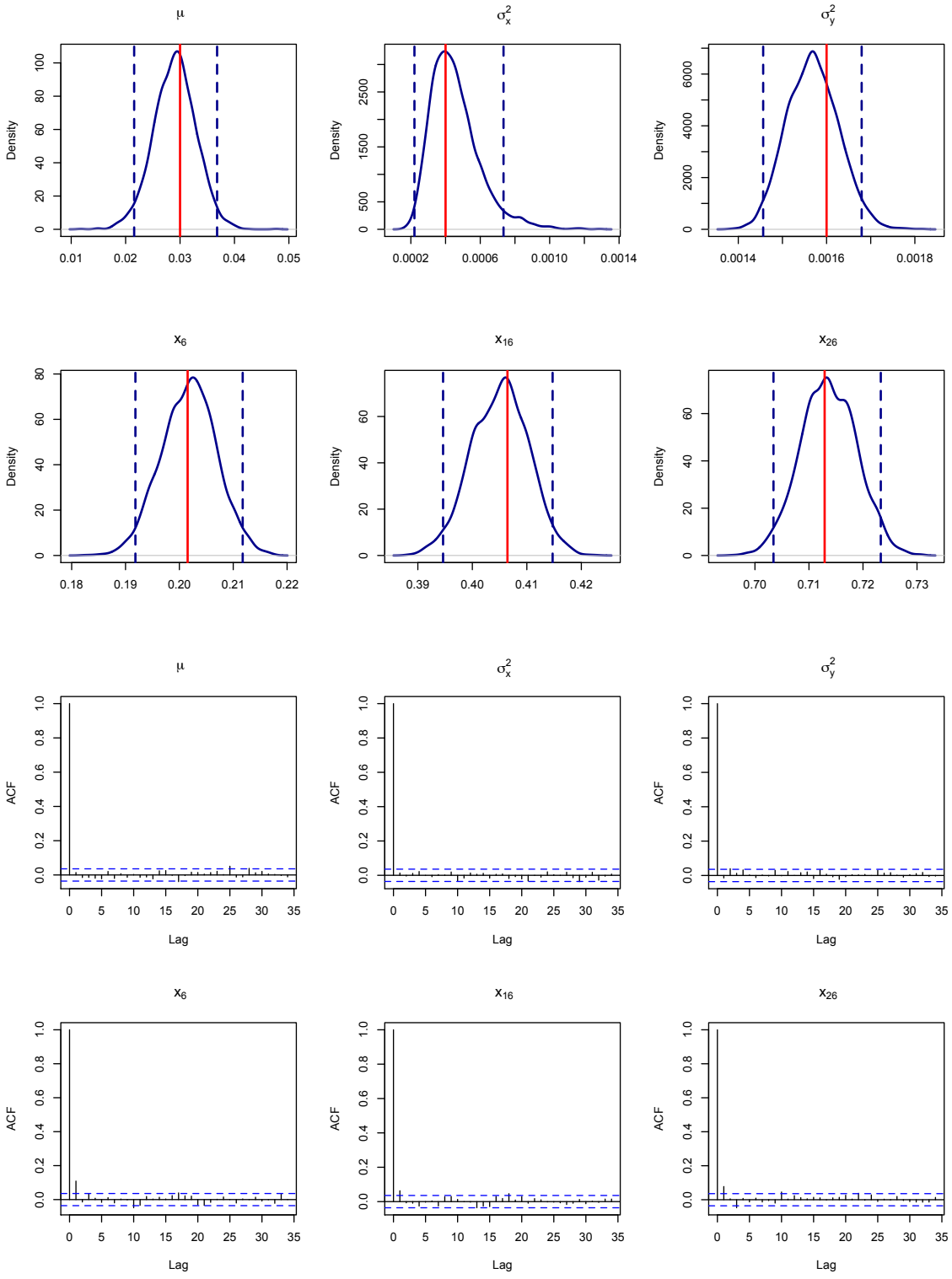

Figure 2.2: Upper two panels: Posterior densities for parameters μ, σ_x^2 and σ_y^2 and for latent states x_6, x_{16} and x_{26} of the Gaussian State Space model. Dashed lines indicate the 95% HPD bands and the red line shows original values used for the simulation. Lower two panels: Autocorrelation functions for parameters μ, σ_x^2 and σ_y^2 and for latent states x_6, x_{16} and x_{26} of the Gaussian State Space model.

significant amount of cross-correlation between the marginal chains of the examined parameters.

As a final point to the exercise, we examine how accurately the full random walk, $\boldsymbol{x}$, has been estimated. Figure 2.4 shows the simulated states $\boldsymbol{x}$ in dots, the estimated means along the solid line and the 95% credibility intervals in dashes. The plot demonstrates the quality of the algorithm since only very few of the states $\boldsymbol{x}$ are out of the constructed intervals.

To summarise, the Gaussian State Space model considered in this section combines the dynamics of two univariate random processes in a hierarchical structure. Such models are also known as hidden Markov models due to their Markov property. A simplification of the model might be found in Tsay (2005). The model was simulated and then estimated under the Bayesian paradigm via a hybrid component-wise MCMC algorithm. The method performs very efficiently, both in terms of estimation precision and posterior sample quality, for all the parameters and latent states involved.

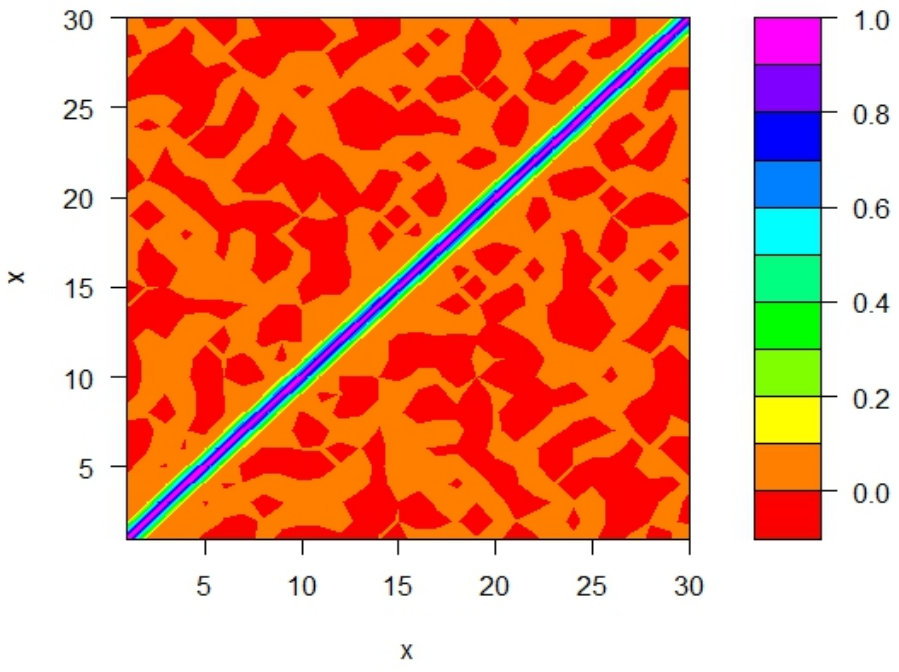

Figure 2.3: Contours of the cross-correlation matrix for the vector of latent states, $\boldsymbol{x}$, of the Gaussian State Space model.

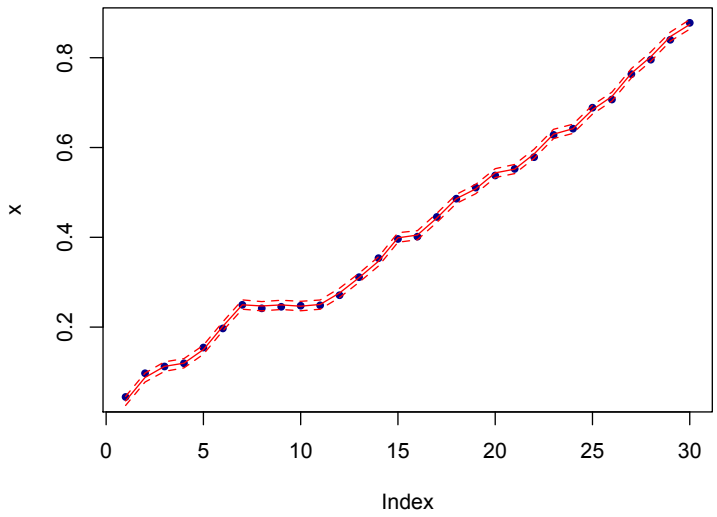

Figure 2.4: Plot of simulated values of the vector $\boldsymbol{x}$ in dots and of the posterior means and 95% credibility intervals from the MCMC sample in solid and dashed lines, respectively.

2.4 Parameter Uncertainty in Mortality Modelling

The parameters of any statistical model are subject to some degree of associated uncertainty. In the context of time-series mortality models the issue of gathering data for longer than, say, 50 years is questionable. Even if the data that far back in time were relevant, it is unclear whether they are relevant to modern times (Sweeting, 2011). With so few observation years available, the parameter estimates of the underlying stochastic model for the mortality dynamics are likely to be misleading. To face the matter of that additional layer of risk in the mortality modelling context researchers have used bootstrap techniques, e.g. as in Brouhns et al. (2005), Renshaw and Haberman (2008) and Li (2010), or resorted to the posterior distribution of the parameters (Cairns et al., 2006a). Plat (2009) refers to all possible ways of incorporating parameter uncertainty and concludes that bootstrapping is the preferred option for practical applications.

The most formal way of incorporating parameter risk is under the Bayesian framework. To that end, Pedroza (2006) implements a Bayesian version of the LC model under the original homescedastic normally distributed errors assumption by sampling through linear Kalman filters. Sampling for the latent states therein is achieved through linear Kalman filters whilst the conjugate properties of the parameters are also employed. The model is estimated via MCMC and the forecasting through sam-

pling the posterior predictive distribution of the rates ensures the incorporation of the underlying parameter risk. Similarly, Czado et al. (2005) suggest another version of the Bayesian LC model with a mean-reverting process around a linear trend for the stochastic factor of the model. Nevertheless, the general process of modelling different sources of uncertainty in an actuarial context is thoroughly discussed by Cairns (2000). Additionally, Cairns et al. (2006a) simulate the CBD model under parameter uncertainty by considering their posterior distributions of the parameters of the associated random walk model. More generally, Bayesian methods have also been applied in two population modelling, under the Age-Period-Cohort framework (Cairns et al., 2011b).

The mortality models projected in Chapter 1 assumed their parameters unknown but fixed, so that no risk was incorporated into their estimates. In this section we employ the Bayesian framework, described so far, to take into account the risk underlying the full parameter vector of the CBD model. That is, we now examine the parameters of the CBD model according to their posterior distributions after some prior specifications are assigned for them.

2.4.1 Bayesian CBD Model

The derivation of the posterior distribution of the CBD model follows from the results developed in the previous Sections. The presented modification of the model shall be referred to as B-CBD. The simple structure of the model makes it easy to introduce prior beliefs, so that the relevant posteriors can be derived. The full description of the model, given in Section 1.3.1, is repeated below together with all accompanying assumptions:

$$\begin{aligned} D(x,t)|m(x,t) &\sim Poi\big(m(x,t)E(x,t)\big) \\ \operatorname{logit}\big(q(x,t)\big) &= \kappa_t^{(1)} + \kappa_t^{(2)}\big(x-\bar{x}\big) \\ \boldsymbol{\kappa}_t|\boldsymbol{\kappa}_{t-1},\boldsymbol{\delta},V_\zeta &\sim N\big(\boldsymbol{\kappa}_{t-1}+\boldsymbol{\delta},V_\zeta\big). \end{aligned}$$

The conditionally normal distribution for the period effect vectors, $\boldsymbol{\kappa}_t$, follows from the random walk model employed, as described in Section 1.4.2. The mortality rates $q(x,t)$ are connected to the central death rates $m(x,t)$ of the likelihood function through the usual relationship given in equation (1.1.7). In order to let the data reveal as much information they carry as possible, all the prior specifications are relatively vague. Hence, the drift vector $\boldsymbol{\delta}$ of the random walk is assigned a bivariate normal prior with mean $\boldsymbol{\delta}_0 = (0,0)$ and covariance matrix $V_0 = I_2$, the 2×2 diagonal matrix. The covariance matrix V_ζ is assigned the uninformative Jeffreys prior, introduced in

Section 2.1.2, with density of the form:

$$p\big(V_\zeta\big) \propto |V_\zeta|^{-3/2}.$$

Therefore, the joint log prior distribution of the vector $\big(\boldsymbol{\delta}, V_\zeta\big)$, whose components are assumed independent among them, is given as:

$$p\big(\boldsymbol{\delta}, V_\zeta\big) \propto \big(\boldsymbol{\delta} - \boldsymbol{\delta}_0\big)' V_0^{-1}\big(\boldsymbol{\delta} - \boldsymbol{\delta}_0\big) - \frac{3}{2}\log\big(|V_\zeta|\big) = \sum_{i=1}^{2} \delta_i^2 - \frac{3}{2}\log\big(|V_\zeta|\big).$$

Now let the global parameter vector of the model be $\boldsymbol{\theta}$, and indicate $\boldsymbol{\theta}_{-i}$ the same parameter vector except those in focus for each respective case examined. The conjugate properties of the particular prior specifications lead to the following full conditional posteriors for the parameters of the random walk of the B-CBD model.

$$V_\zeta|\boldsymbol{\theta}_{-i} \;\sim\; IW\Big(n-2, \frac{1}{n-1}\widehat{V}\Big) \tag{2.4.1}$$

$$\boldsymbol{\delta}|\boldsymbol{\theta}_{-i} \;\sim\; N_2\big(\boldsymbol{\delta}_p, V_p\big), \tag{2.4.2}$$

with

$$\begin{aligned}
\widehat{\boldsymbol{\delta}} &= \frac{1}{n-1}\sum_{i=1}^{n-1}\big(\boldsymbol{\kappa}_{i+1} - \boldsymbol{\kappa}_i\big) \\
\widehat{V} &= \frac{1}{n-1}\sum_{i=1}^{n-1}\left\{\Big(\boldsymbol{\kappa}_{i+1} - \boldsymbol{\kappa}_i - \widehat{\boldsymbol{\delta}}\Big)\Big(\boldsymbol{\kappa}_{i+1} - \boldsymbol{\kappa}_i - \widehat{\boldsymbol{\delta}}\Big)'\right\} \\
\boldsymbol{\delta}_p &= \Big(V_0^{-1} + (n-1)V_\kappa^{-1}\Big)^{-1}\Big(V_0^{-1}\boldsymbol{\delta}_0 + (n-1)V_\kappa^{-1}\widehat{\boldsymbol{\delta}}\Big) \\
V_p &= \Big(V_0^{-1} + (n-1)V_\kappa^{-1}\Big)^{-1},
\end{aligned} \tag{2.4.3}$$

where $IW(f, S)$ denotes the Inverse Wishart distribution with f degrees of freedom and scale matrix S, which is the commonly employed conjugate prior for the covariance matrix of the Normal distribution (Gelman et al., 2003).

The full conditional posteriors for the vectors $\boldsymbol{\kappa}_t$ do not possess any known form of distribution. Instead, the interplay of the Poisson assumption with the time-series structure builds the mixture they consist of. More particularly, the densities for the first and last vectors, $\boldsymbol{\kappa}_1$ and $\boldsymbol{\kappa}_n$ respectively, contain a single term from the normal density. In contrast, the intermediate terms, $\boldsymbol{\kappa}_j$ include two such terms from the normal density. We can define the likelihood terms for each year t, as a function of

the vector of period effects of the model, as follows:

$$\ell\left(\boldsymbol{\kappa}_t\right) = \sum_x \left\{ D(x,t) \log\left\{ \log\left[1+\exp\left(\phi(\boldsymbol{\kappa}_t)\right)\right]\right\} - E(x,t)\log\left[1+\exp\left(\phi(\boldsymbol{\kappa}_t)\right)\right]\right\}.$$

Function ϕ determines the parameterisation of the CBD model, which only depends on the vector of period effects for fixed calendar year t. Thus, each of the terms just defined expresses the contribution of the Poisson likelihood in the model for year t.

The full conditional posteriors of the $\boldsymbol{\kappa}_t$ vectors take the following forms, for the first, intermediate and last year used in the estimation, respectively:

$$\begin{aligned}
\pi\left(\boldsymbol{\kappa}_1|\boldsymbol{\theta}_{-i}\right) &= \ell\left(\boldsymbol{\kappa}_1\right) - \frac{1}{2}\Big[\left(\boldsymbol{\kappa}_2-\boldsymbol{\kappa}_1-\boldsymbol{\delta}\right)' V_\kappa^{-1}\left(\boldsymbol{\kappa}_2-\boldsymbol{\kappa}_1-\boldsymbol{\delta}\right)\Big], \\
\pi\left(\boldsymbol{\kappa}_j|\boldsymbol{\theta}_{-i}\right) &= \ell\left(\boldsymbol{\kappa}_j\right) - \frac{1}{2}\Big[\left(\boldsymbol{\kappa}_{j+1}-\boldsymbol{\kappa}_j-\boldsymbol{\delta}\right)' V_\kappa^{-1}\left(\boldsymbol{\kappa}_{j+1}-\boldsymbol{\kappa}_j-\boldsymbol{\delta}\right) \\
&\qquad + \left(\boldsymbol{\kappa}_j-\boldsymbol{\kappa}_{j-1}-\boldsymbol{\delta}\right)' V_\kappa^{-1}\left(\boldsymbol{\kappa}_j-\boldsymbol{\kappa}_{j-1}-\boldsymbol{\delta}\right)\Big], \text{with } 1<j<n, \\
\pi\left(\boldsymbol{\kappa}_n|\boldsymbol{\theta}_{-i}\right) &= \ell\left(\boldsymbol{\kappa}_n\right) - \frac{1}{2}\Big[\left(\boldsymbol{\kappa}_n-\boldsymbol{\kappa}_{n-1}-\boldsymbol{\delta}\right)' V_\kappa^{-1}\left(\boldsymbol{\kappa}_n-\boldsymbol{\kappa}_{n-1}-\boldsymbol{\delta}\right)\Big].
\end{aligned}$$

As the vector of period effects, $\boldsymbol{\kappa}_t$ for fixed year t, is implied to have correlated components, they are sampled jointly for a fixed year. A classical MH step is used in the algorithm leading to a hybrid MCMC scheme for estimating the B-CBD model. If the current state of the chain is $\boldsymbol{\kappa}_t^*$ for some year t, the posterior covariance matrix of the vector $\boldsymbol{\kappa}_t$ may be approximated by the inverse of the Observed Information Matrix, which is given as:

$$\mathcal{I}(\boldsymbol{\kappa}_t^*) = -\left(\begin{array}{cc} \frac{\partial^2 \pi(\boldsymbol{\kappa}_t|\boldsymbol{\theta}_{-i})}{\left(\partial \kappa_t^{(1)}\right)^2} & \frac{\partial^2 \pi(\boldsymbol{\kappa}_t|\boldsymbol{\theta}_{-i})}{\left(\partial \kappa_t^{(1)}\right)\left(\partial \kappa_t^{(2)}\right)} \\ \frac{\partial^2 \pi(\boldsymbol{\kappa}_t|\boldsymbol{\theta}_{-i})}{\left(\partial \kappa_t^{(1)}\right)\left(\partial \kappa_t^{(2)}\right)} & \frac{\partial^2 \pi(\boldsymbol{\kappa}_t|\boldsymbol{\theta}_{-i})}{\left(\partial \kappa_t^{(2)}\right)^2} \end{array} \right) \Bigg| \boldsymbol{\kappa}_t^*. \tag{2.4.4}$$

The calculation of the above approximation is simplified by the mixture of Poisson and quadratic forms of the conditional log-posteriors for $\boldsymbol{\kappa}_t$. The inverse of $\mathcal{I}(\boldsymbol{\kappa}_t^*)$ is multiplied by some constant c to improve the acceptance rates of the $\boldsymbol{\kappa}_t$ vectors. In the particular implementation, c was set equal to 2.4. Summarising, values of the $\boldsymbol{\kappa}_t$ are sampled according to:

$$\boldsymbol{\kappa}_t \sim N_2\left(\boldsymbol{\kappa}_t^*, c\mathcal{I}^{-1}(\boldsymbol{\kappa}_t^*)\right). \tag{2.4.5}$$

The above developed hybrid MCMC shceme is summarised in the following pseudo-algorithm.

Algorithm 3

Initialise the parameter vector

$$\boldsymbol{\theta}^{(0)} = \left\{ \boldsymbol{\kappa}_t^{(0)}, \boldsymbol{\delta}^{(0)}, V_\zeta^{(0)} \right\},$$

and set the number of iterations, M.
For $j = 1, \ldots, M$

- For $i = 1, \ldots, n$
  - Sample a candidate period effects vector, $\boldsymbol{\kappa}_i^*$, from the proposal distribution of equation (2.4.1), with density $q_{\boldsymbol{\kappa}_i}$.
  - Calculate the acceptance ratio according to:

$$r\left(\boldsymbol{\kappa}_i^{(j-1)}, \boldsymbol{\kappa}_i^*\right) = min\left\{1, \frac{q_{\boldsymbol{\kappa}_i}\left(\boldsymbol{\kappa}_i^{(j-1)} | \boldsymbol{\kappa}_i^*\right) \times \exp\left(\pi\left(\boldsymbol{\kappa}_i^* | \boldsymbol{\theta}_{-i}\right)\right)}{q_{\boldsymbol{\kappa}_i}\left(\boldsymbol{\kappa}_i^* | \boldsymbol{\kappa}_i^{(j-1)}\right) \times \exp\left(\pi\left(\boldsymbol{\kappa}_i^{(j-1)} | \boldsymbol{\theta}_{-i}\right)\right)}\right\}.$$

  - Generate $U \sim U(0,1)$.
  - If $U \leq r\left(\boldsymbol{\kappa}_i^{(j-1)}, \boldsymbol{\kappa}_i^*\right)$, set $\boldsymbol{\kappa}_i^{(j)} = \boldsymbol{\kappa}_i^*$, otherwise set $\boldsymbol{\kappa}_i^{(j)} = \boldsymbol{\kappa}_i^{(j-1)}$.
- Sample the drift $\boldsymbol{\delta}^{(j)}$, and the covariance matrix $V_\zeta^{(j)}$, from their full conditional posterior distributions given in the set of equations (2.4.1).
- Save the iteration $\boldsymbol{\theta}^{(j)}$ and continue if $j < M$.

2.4.2 Convergence & Posterior Inference for the Bayesian CBD Model

The model is calibrated to the EW males data-set and the results are compared to those of the Poisson MLE solution of Section 1.3. The algorithm runs for 100,000 iterations, and after discarding the initial 10% of them as burn-in period, only every 30^{th} iteration is kept to reduce the autocorrelation of the chains for the estimated parameters. Hence, we end up with a posterior sample of size around 3,000.

The convergence is assessed with the diagnostics developed in Section 2.2.3. The marginal MCMC samples are tested with Geweke's diagnostic by comparing the mean of the first 20% of the convergent chain to that of the final 40%. Table 2.3 reports the resultant Z-scores and corresponding p-values for the parameters of the bivariate random walk model, $\boldsymbol{\delta}$ and V_ζ, for the period effects, $\boldsymbol{\kappa}_t$. Except for the drift δ_1 all the other parameters do not contradict the stationary mean hypothesis at any salient significance level. The identical test is employed for all the states of the period effects,

$\boldsymbol{\kappa}_t$. Out of one hundred tests of all the individual chains, only 7 times the underlying hypothesis could be rejected at a 10% significance level, and out of these, three cases would also not stand at a 5% level.

	δ_1	δ_2	σ_1^2	σ_2^2	σ_{12}
Z-score	-2.3673	0.1927	-1.083	-0.408	-0.6256
p-value	0.0179	0.8472	0.2788	0.6832	0.5316

Table 2.3: Z-scores and p-values from Geweke's diagnostic for parameters of the random walk, $\boldsymbol{\delta}$ and V_ζ, of the B-CBD model, EW data.

The convergence of the algorithm is also assessed visually through the cumulative trace-plots of the posterior means and 95% HPD intervals. The graphs of Figure 2.5 also show, where applicable and in red dashed lines, the MLE's from the Poisson implementation of the CBD model in Section 1.3.2. The posterior distributions of the random walk parameters, $\boldsymbol{\delta}$ and V_ζ, generally seem to be in agreement with the Poisson MLE's. The variance of $\kappa_t^{(2)}$, σ_2^2, appears as the only case in which the mean posterior estimate is concentrated slightly lower than the corresponding MLE. However, the deviation is very small in absolute value, and no significant difference should occur in the projections under the two different implementations of the model due to that reason. On the other hand, as it will be shown later on, the posterior means of the $\boldsymbol{\kappa}_{50}$ terms exhibit notable shifting relative to the MLE's, so that they are not shown in the respective cumulative plots. The deviation of the Bayesian estimates compared to the MLE's is more frequent for $\kappa_t^{(2)}$ and only apparent for the latest years' $\kappa_t^{(1)}$.

The graphs of δ_1 and $\kappa_{50}^{(2)}$ in Figure 2.5 assist in identifying the reason for the low p-value produced by Geweke's diagnostic. According to the cumulative trace-plots, during the first few hundred iterations of the chain their values oscillate strongly relative to the final convergence point. If the cumulative plots are volatile then the successive iterations of the chain are far apart and the parameter is not being efficiently estimated. However, the long term behaviour of all the chains is satisfying, regardless of the negative result of the diagnostic.

Table 2.4 summarises the posterior distributions of the parameters of the random walk model and of the model's parameter estimates for the last observation year, $\boldsymbol{\kappa}_{2009}$. Beyond the posterior means and standard deviations of the distributions, the 95% HPD intervals are also tabulated. The acceptance rate for the MH steps of the scheme are between 21-27% across all applicable parameters. The presented distributions will be used to project the model under the correct degree of underlying uncertainty. The variability of the Bayesian estimates is also assessed via the Coefficient of Variation

(CV) of their posterior distributions. The CV is defined as the absolute quotient of the standard deviation to the mean of the posterior distribution in percentage form, and will also be used elsewhere in the Thesis to assist in adjusting the spread of prior distributions. The higher the percent of the CV, the higher the variability of the distribution, and the figure might well exceed 100%.

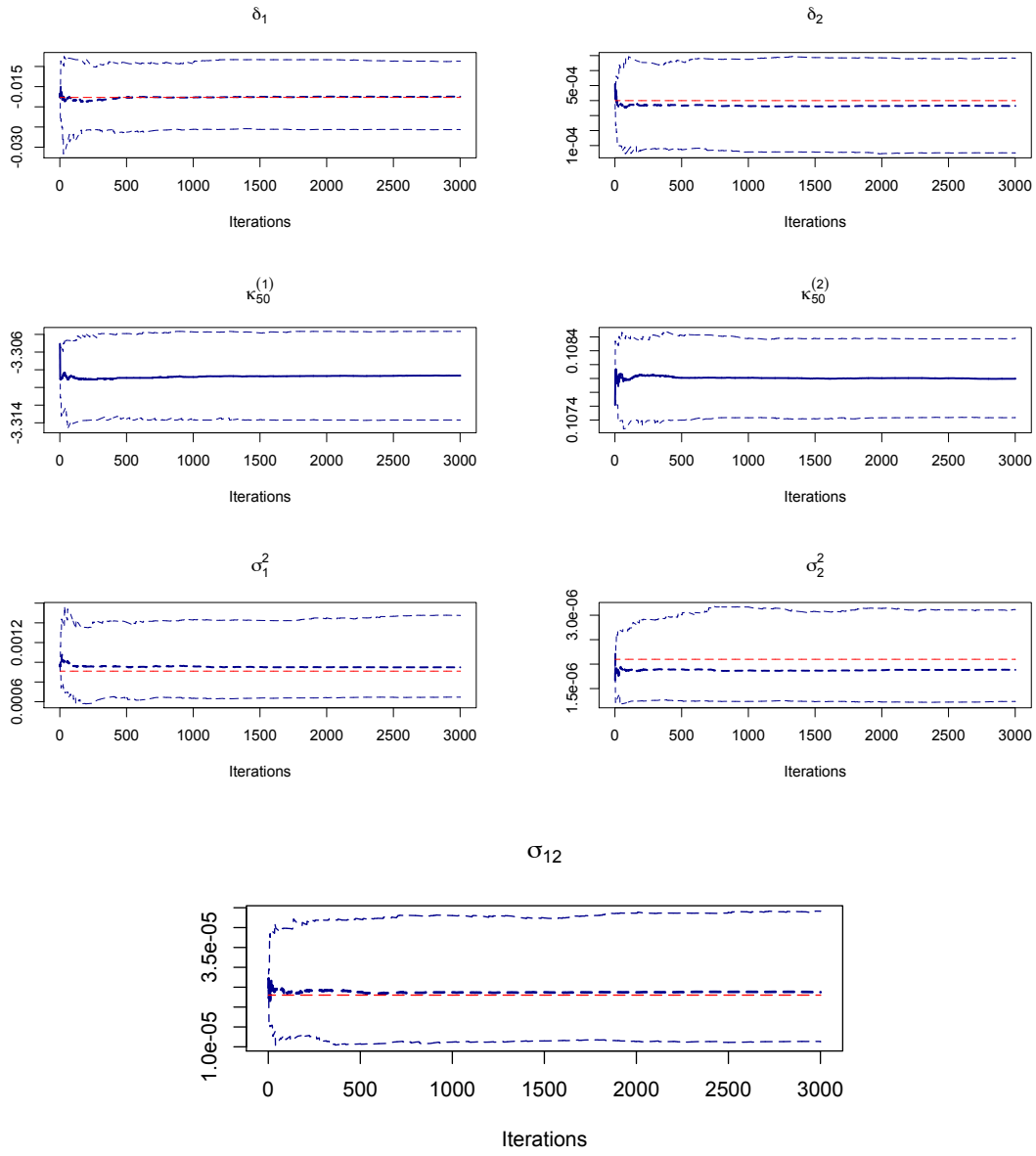

Figure 2.5: Cumulative trace-plots of mean and 95% HPD intervals for parameters of the bivariate random walk, $\boldsymbol{\delta}$ and V_ζ, and of the vector of period effects, $\boldsymbol{\kappa}_{2009}$, of the B-CBD model, EW data.

The distribution of δ_2 has CV 45% and that of σ_{12} has 32%. The average statistic over all five posteriors of the random walk parameters, $\boldsymbol{\delta}$ and V_ζ, is 30%. Parameters $\kappa_t^{(1)}$ exhibit much lower CV with range between 0.07% and 0.1%, whereas the

coefficient is between 27% and 33% for $\kappa_t^{(2)}$. Notably, for both sets of parameters the coefficients are of descending order with calendar year. Hence, the latent states for the mortality risk factors are more accurate for later years of the data-set. The above amount of variability imposed by the Bayesian solution is supplementing the CBD model with the proper degree of uncertainty regarding all the parameters and latent states involved.

	Mean (MLE)	Std Deviation	L_{hpd}	U_{hpd}
δ_1	$-0.01737(-0.01767)$	4.51×10^{-3}	-0.0266	-0.0087
δ_2	$3.633 \times 10^{-4}(3.997 \times 10^{-4})$	1.638×10^{-4}	6.24×10^{-5}	6.93×10^{-4}
σ_1^2	$9.682 \times 10^{-4}(9.102 \times 10^{-4})$	2.103×10^{-4}	6.067×10^{-4}	1.37×10^{-3}
σ_2^2	$1.961 \times 10^{-6}(2.098 \times 10^{-6})$	4.809×10^{-7}	1.169×10^{-6}	2.95×10^{-6}
σ_{12}	$2.458 \times 10^{-5}(2.301 \times 10^{-5})$	8.068×10^{-6}	1.043×10^{-5}	4.152×10^{-5}
$\kappa_{2009}^{(1)}$	$-3.309(-3.321)$	2.516×10^{-3}	-3.313	-3.303
$\kappa_{2009}^{(2)}$	0.108 (0.110)	2.937×10^{-4}	0.1074	0.1086

Table 2.4: Posterior summary of the B-CBD model, EW data.

The quality of the posterior sample may be examined as discussed in Sections 2.3 and 2.2.3. Figure 2.6 shows the ACF's of the marginal chains of the random walk parameters, $\boldsymbol{\delta}$ and V_ζ. Along with the vector $\boldsymbol{\kappa}_{2009}$, these parameters are essential for the forecasts of future rates. Their ACF's show no statistically signifiant amount of autocorrelation at any lag.

The cross-correlation between the parameters of interest is examined through the correlation matrix of their marginal chains. For the parameter vector in Table 2.4, $\left(\delta_1, \delta_2, \sigma_1^2, \sigma_2^2, \sigma_{12}, \kappa_{2009}^{(1)}, \kappa_{2009}^{(2)}\right)$, the corresponding matrix is given in Table 2.5. No significant amount of correlation is detected within $\boldsymbol{\delta}$ or among the entries of $\boldsymbol{\delta}$ and any other parameters. The elements of the covariance matrix, V_ζ, appear strongly correlated, especially σ^{12} with σ_1^2 and σ_2^2, respectively. The marginal chains of σ_1^2, σ_2^2 and σ_{12} are obtained jointly via sampling the IW draw. The root of the dependence structure between the elements of V_ζ is due to the sampling method for that matrix. Alternatively, one could consider the individual full conditional posteriors of those parameters and implement an algorithm which samples individually for each one of them. Obviously, the IG distribution would be the choice for the prior distribution, and the structure of the model yields conjugate full conditional posteriors. The elements of V_ζ do not appear correlated with any other entries of the parameter sub-vector considered here. Finally, the vector of period effects $\boldsymbol{\kappa}_{2009}$ also produces negative correlation. However, these are sampled jointly from a normal approximation

to their posterior, although under a MH step, and thus the result was first, anticipated and second, negligible.

	δ_1	δ_2	σ_1^2	σ_2^2	σ_{12}	$\kappa_{2009}^{(1)}$	$\kappa_{2009}^{(2)}$
δ_1	1						
δ_2	0.0418	1					
σ_1^2	0.01062	0.0335	1				
σ_2^2	0.0023	-0.0283	0.3307	1			
σ_{12}	-0.0091	-0.004	0.6924	0.7030	1		
$\kappa_{2009}^{(1)}$	-0.013	0.0028	-0.0014	0.0087	0.0469	1	
$\kappa_{2009}^{(2)}$	0.0018	0.0156	0.0128	0.0249	-0.0215	-0.1288	1

Table 2.5: Cross-correlation matrix of the MCMC series for the parameter sub-vector $\left(\delta_1, \delta_2, \sigma_1^2, \sigma_2^2, \sigma_{12}, \kappa_{2009}^{(1)}, \kappa_{2009}^{(2)}\right)$ of the B-CBD model, EW data.

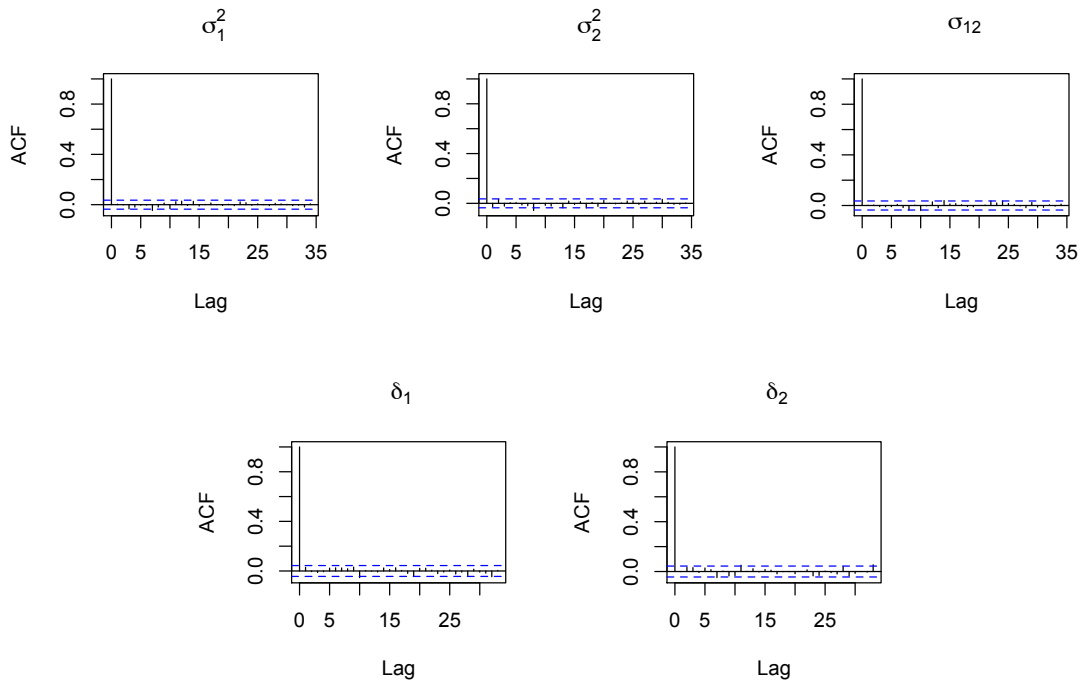

Figure 2.6: Autocorrelation functions of the MCMC sample for the drift vector, $\boldsymbol{\delta}$, and the elements of the covariance matrix, V_ζ, of the B-CBD model, EW data.

The full set of the posterior means for the latent state estimates, $\boldsymbol{\kappa}_t$, of the B-CBD model is shown in blue dots in Figure 2.7. The red circles indicate the Poisson MLE's of the original CBD model. The plot shows some deviation in the posterior mean estimates of the $\kappa_t^{(2)}$ set relative to the MLE's of the Poisson model. Similar deviation

is observed only for the latest years of observation for the $\kappa_t^{(1)}$ series. The distortion of the slope parameters however, is not enough to affect significantly the fitted rates. Since the main pattern of the mortality profile is determined by the level parameters $\kappa_t^{(1)}$, only the deviations of those in the latest years of observation could be capable of significantly tilting the fitted death rates. On the other hand, plotting the fitted rates, and assuming that the posterior mean is the point representative estimate of the B-CBD model, the difference between the two models is indistinguishable. Furthermore, the credible intervals are narrow enough, so that the difference between the two models in terms of goodness-of-fit for up to 95% posterior probability is negligible.

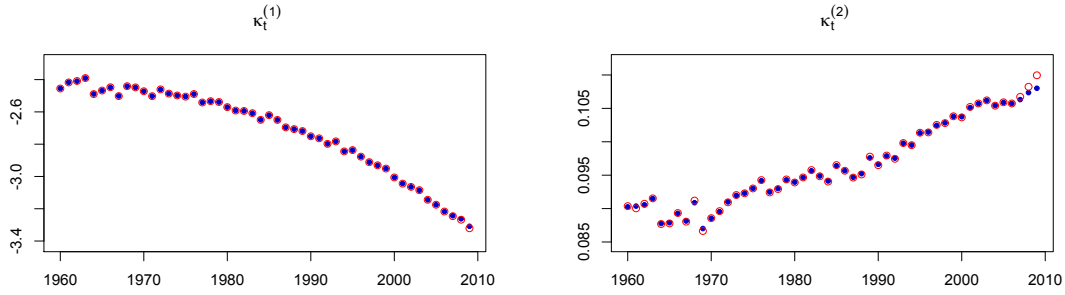

Figure 2.7: Posterior means of the period effects, $\boldsymbol{\kappa}_t$, of the B-CBD model in blue dots along with the MLE's of the original CBD model under the Poisson model in red circles, EW data.

2.4.3 Comparative Simulations and Forecasts

The posterior distribution of the B-CBD model is used to forecast future rates and relevant mortality indices and statistics. The model is firstly projected by assuming the involved latent states and parameters are fixed unknowns, whose estimates are taken as their respective posterior means. The method is said to be Parameter Certain (PC) since no source of parameter randomness is taken into account. Alternatively, sufficiently many random draws from the posterior sample are used to get different parameter vectors from the MCMC sample, and for each one of them project the future rates under Parameter Uncertainty (PU). Both methods yield predictive distributions for the future rates. The latter also incorporates the proper degree of uncertainty underlying the latent states and parameters of the CBD model. Finally, the probability of survivorship is followed for individuals aged 65 in 2010 for the following 25 years, under both methods previously described. All the results are compared to those of the Poisson implementation of the CBD model.

Firstly, to simulate the model under PC, denote by $\boldsymbol{\theta}$ the full parameter vector of the CBD model. The MCMC sample consists of $\boldsymbol{\theta}_i$, $i = 1, \ldots, M$ such vectors, where M is the size of the posterior sample. The posterior mean of the parameter vector is:

$$\widehat{\boldsymbol{\theta}} = \frac{1}{M} \sum_{i=1}^{M} \boldsymbol{\theta}_i.$$

The model is projected for $Y = 50$ years ahead and $N = 2,000$ iterations. First, the random walk is forecasted and then the estimates are used to construct the future rates. The random walk starts from the posterior mean estimate $\widehat{\boldsymbol{\kappa}}_{2009}$ of the vector $\widehat{\boldsymbol{\theta}}$. To construct the required rates we also use the function ϕ, defined in Section 2.4.1, which gives the parameterisation of the logit mortality rates in terms of the period effects and the fixed age mortality basis.

The algorithm for the projection is as follows:

- for $j = 1, \ldots, N$
  - for $s = 1, \ldots, Y$
    $\widehat{\boldsymbol{\kappa}}_{2009+s} = \widehat{\boldsymbol{\kappa}}_{2009+s-1} + \widehat{\boldsymbol{\delta}} + \widehat{C}_\zeta \boldsymbol{\omega}_s$
    get the vector of future rates for year $2009 + s$ as
    $\widehat{\boldsymbol{q}}_{2009+s} = \text{logit}^{-1}\Big(\phi\left(\widehat{\boldsymbol{\kappa}}_{2009+s}\right)\Big)$
  - end for;

  store the simulated matrix, end for;

The vectors $\boldsymbol{\omega}_s$ are *i.i.d.* standard bivariate normal variables for $s = 1, \ldots, Y$ and the matrix $\widehat{C}_\zeta$ is derived from the Cholesky decomposition of the estimate $\widehat{V}_\zeta$, taken from the global vector $\widehat{\boldsymbol{\theta}}$. The method yields an array of matrices, from which we get the predictive distributions of the future rates.

Secondly, to simulate the model under PU, j indices with $j = 1, \ldots, N$ are randomly drawn, with range from 1 to the size of our posterior sample, M. For each set of parameters corresponding to these indices, say $\boldsymbol{\delta}^{(j)}$ and $C_\zeta^{(j)}$, the model is projected to return future rates for Y years ahead. This again yields posterior predictive distributions of the future rates, but since randomly chosen parameter sets are used from the full posterior sample instead of the mean estimates of the PC case, the method incorporates all the underlying parameter uncertainty of the model.

The algorithm for the simulation follows:

- for $j = 1, \ldots, N$
  - for $s = 1, \ldots, Y$

$\boldsymbol{\kappa}^{(j)}_{2009+s} = \boldsymbol{\kappa}^{(j)}_{2009+s-1} + \boldsymbol{\delta}^{(j)} + C^{(j)}_{\zeta}\boldsymbol{\omega}_s$

get the vector of future rates for year $2009+s$ as

$\widehat{\boldsymbol{q}}_{2009+s} = \text{logit}^{-1}\left(\phi\left(\boldsymbol{\kappa}^{(j)}_{2009+s}\right)\right)$

– end for;

store the simulated matrix, end for;

In this case, the random walk starts up at the estimates of $\boldsymbol{\kappa}^{(j)}_{2009}, j = 1, \ldots, N$ included in the MCMC sample $\boldsymbol{\theta}_i, i = 1, \ldots, M$.

The above developed projections of the B-CBD model are compared to those of the Poisson CBD model. There are only slight differences between the expected mean forecast of the PC B-CBD forecasts and that of the Poisson CBD, which are apparent only for younger ages. This might be a consequence of the improved fitting of the Bayesian model to the realised rates, due to the deviation in the estimates of the random walk states for the latest years of the dataset. Even in this case, the observed differences take effect only at least 30 years ahead of the start of the simulation, that is, around 2040. For ages older than around 70, the differences in the mean forecasts between the two projections are minimal and hardly recognisable. Generally, the confidence bands of the PC B-CBD projection are usually very close to those of the Poisson CBD model. The risk of the parameters of the CBD model, imposed through the Bayesian implementation and fully incorporated under the second simulation scheme, turns out to introduce significant differences only in the long term of the projection. In particular, the older the age examined, the shorter the period in which PU becomes significant and produces notably wider confidence bands. For example, PU yields distinguishable differences at around 10 years ahead of the start of the forecast for (85), but at 20 years ahead when considering rates of (65). This in turn results in proportionally greater confidence bands for the older ages of the dataset. In all examined cases, the PU B-CBD projection overlaps all existing confidence bands in the examined plots.

From the projected predictive distributions of mortality rates, we also extract the distribution of the survivor index for males (65) in 2010, $S(65, t)$. The index returns the proportion of the population surviving at each subsequent year of simulation, it was developed in equation (1.1.1) and was also used in Section 1.4.3. The specific cohort is followed for the next 25 years and therefore, we obtain the predictive distribution of the probability of surviving for the examined cohort until they reach age 89 in year 2034. The described distribution is extracted for both the PC and PU projection methods, and compared to the results of the Poisson MLE solution of the CBD model. The graphical illustrations of the above survivor distributions are almost identical to the graphs of figure 1.12, and since deviations between the examined

models are minimal, are excluded.

There are very few differences when contrasting the simulated survivor indices of the three different projections of the CBD model. Most notably, the parameter uncertainty becomes visually distinguishable beyond the first 15 years of the projection period. As a means of example we examine the index $S(65,t)$ for $t = 20$ for all three models in Table 2.6. Beyond the mean of its distribution, the table includes its extrema and the percentiles, denoted as $l_{2.5\%}$ and $u_{97.5\%}$, required for the 95% confidence bands. The projection of the MLE solution produces the higher mean probability of survivorship, but the PU projection of the B-CBD model overlaps all other confidence intervals, and also produces the most extreme values for $S(65,20)$. However, the overall differences between the three projections are minimal. Hence, the quantities and indices based on $S(65,t)$ do not produce any particularly distinctive results. Indeed, differences in either the period or cohort life-expectancy distributions for the original CBD and the PC B-CBD models are observed only at relatively old ages and extended projection horizons. Specifically, the PC B-CBD model produces markedly higher means and variances than the Poisson CBD model, for ages of 75 and above and simulation horizons of 30 years or more. As expected, the PU Bayesian forecasts yield distributions which have distinguishably greater spread, but only as the simulation horizon prolongs significantly.

This section provides a Bayesian implementation of the CBD model using a hybrid MCCM algorithm. The developed stochastic simulation scheme will be used as the base to estimate the suggested extension of the CBD model in Chapter 5. The vague prior information supplied in the algorithm yield a posterior distribution which is mainly determined by the information contained in the data. The results were compared to those derived from the Poisson implementation of the CBD model, given in Section 1.3.2. Projections were conducted both under PU and PC regimes and we conclude that first, Bayesian PC forecasts have little difference to the MLE Poisson solution and second, that parameter uncertainty becomes significant in the medium to the long term of the projections. Finally, the importance of projecting under PU for medium and long term predictions was verified by calculating the relevant mortality indices for indicative ages across the whole population.

$S(65,20)$	Mean	Min	Max	$l_{2.5\%}$	$u_{97.5\%}$
CBD	.5023	.3895	.6085	.4249	.5751
B-CBD, PC	.4999	.3835	.6087	.4195	.5749
B-CBD, PU	.5010	.3352	.6243	.4156	.5820

Table 2.6: Predictive distributions of survivor indices for EW males aged 65 in 2010 after $t = 20$ years, between the CBD, PC-CBD and PU-CBD projections.

2.5 Summary

This Chapter presented the fundamental elements of Bayesian modelling and the construction of the associated MCMC estimation algorithms. A linear Gaussian State-Space models was used as an example of how a hybrid MCMC algorithm may be developed to estimate the latent states and parameters of the model. The methodology was extended in the context of stochastic mortality models by implementing a Bayesian version of the CBD model for the EW data, and its outcomes were compared to those of the CBD model in Chapter 1. The posterior sample is considered qualitative enough to produce reliable forecasts. The differences under the B-CBD are mostly evident for the latest of the states $\kappa_t^{(2)}$ and for the first component the drift vector, $\boldsymbol{\delta}$. Generally, different parameter sets have different amount of posterior variability, but the underlying dynamics of the two versions of the CBD model are very similar. The imposed uncertainty implied by the Bayesian solution gets prominent mostly as the projection horizon is extended, but the differences in the mean forecasts are generally slight.

Chapter 3

Residuals Modelling

This Chapter provides the motivation and foundations for the extended modelling framework suggested by the Thesis. The estimated mortality residuals of the LC and CBD models from Chapter 1 for the EW data are revisited and assessed relative to their statistical properties and the correlation structures they exhibit. Based on the results, an appropriate vector autoregressive (VAR) process is employed to model the latent residual states in a parsimonious manner. The exact specification of the residuals model depends on the benchmark structure used, and in that sense, the two factors of the CBD model are proven beneficial. The stationarity of the residuals model implies serial correlation between the cells of the population which have the sense of short-term mortality dependencies. Finally, the residuals model is fitted to the outcome of the LC and CBD models from the EW data as a means of verification of the underlying assumptions.

3.1 Introduction

As was shown in Chapter 1 there is an obvious trade-off in the behaviour of stochastic mortality models. The remarkably simple structures, such as the LC or the CBD models, appear unable to capture the full underlying dynamics. Therefore, the models do not satisfy the hypotheses they are based on, though such models might be improved by the incorporation of appropriate further stochastic components, so that the data are fully exploited, as was shown in Sections 1.2.3 and 1.3.3.

Models including cohort effects have been found to improve significantly the fit across a variety of datasets (Renshaw and Haberman, 2006, Plat, 2009). However, the age-cohort interaction seems to lack robustness since, fitting the model to a different range of years can lead to qualitatively different parameter estimates (CMI, 2007b, Dowd et al., 2010a). That is, there is a linear relationship between age, period and cohort effects and thus, cohort effects may be replaced approximately by appropriate age and period effects. The aforementioned drawbacks inevitably are fed into the projections of the future mortality rates and related indices. For example, the forecasts conducted in Section 1.4.2 showed that there is inconsistent behaviour of the projected values when using estimated and simulated cohort factors. Instead of adding further effects into existing simple and transparent stochastic mortality models, we examine their residuals and identify directly structures therein. The recognition of those patterns and the statistical properties of the resultant matrices serve to determine an additional stochastic model, which models the residuals and contributes short-term intercorrelations within the examined datasets.

The assessment of goodness-of-fit for a statistical model consists of three stages: estimation, implementation and testing. The LC and CBD models are estimated as described in Sections 1.3.2 and 1.2.2. Under the current estimation assumption, namely the numbers of deaths are conditionally independent Poisson random variables in each data-cell given the relevant death rates, the distribution of deaths should be well approximated by a normal distribution, as the size of the population and the number of deaths gets larger. In turn, the standardised variables should be approximately standard normally distributed. The construction of the standardised residuals was conducted in Section 1.4.1.

We construct the aforementioned matrices for both the LC and the CBD models for the EW data, which we examine by visual diagnostics, such as those employed in Chapter 2. Furthermore, the residuals matrices are decomposed into $m = 30$ univariate time series, for each age in the dataset. Similarly, we take $n = 50$ more time series corresponding to each year of the data-set. Each of these univariate time series is tested for the prediction of *iid* $N(0, 1)$ observations.

The runs test (Gibbons and Chakraborti, 2011) is used to test the hypothesis for the observations in each series being independent. The runs test is a classic non-parametric diagnostic for identifying whether observations of two different types are randomly drawn from their underlying distribution. For our purpose, the two categories the residuals are divided to are negative and positive signed values. Upon acceptance of the independence assumption, the sample series are tested for normality by employing the Jarque-Bera (JB) test (Jarque and Bera, 1987), based on the sample's skewness and kurtosis of a set of independent sample values. Further, for the instances where the previous assumptions are verified we test the zero mean hypothesis with the two-sided, single sample t-test (Sheskin, 2003), and assess the samples' variances with the relevant two-sided χ^2-test (Sheskin, 2003). Additionally, we test the hypothesis that adjacent series of residuals, in both dimensions, are uncorrelated using the t-test for their Pearson product moment correlation coefficient (Dowd et al., 2010b). Finally, we examine the nature of cross-correlation successive residuals series exhibit through their estimated cross-correlation functions. Details about the calculations of the above tests are given in Appendix B.

3.2 Residual Analysis

The standardised residuals of the LC and the CBD models are analysed after calibrating the models to the EW data-set and serve as a working example for the specification of the residuals process. Once the state variables for each model are obtained, they are substituted in the respective modelling equations to yield the fitted rates, $\widehat{m}(x,t)$, across the two dimensions. The initial analysis of the standardised residuals for the EW data-set conducted in Section 1.4.1 showed that their values are far from satisfying the underlying hypotheses due to their excessive values and spread, but also because of the observed patterns they display.

3.2.1 Visual Diagnostics

Figure 3.1 shows the contours of the estimated residuals matrices from the two models for the data used. The LC model in the left panel reveals strong diagonal clusters of positive and negative residual values, indicating the dependency of mortality rates upon the year of birth which is prominent in the data and not accounted for by the model. In the right panel of the figure, the CBD model indicates larger concentrated regions of residuals with the same sign compared to the LC model. However, the respective values are slightly lower than those of the LC model. Of course, the matrix emphasises again the magnitude of the cohort dependency within the EW data.

The misbehaviour of the residuals matrices is also assessed in Figure 3.2, by plotting

their values against the relevant ages, years and cohorts' years of birth for both models. The LC model develops heteroscedasticity as the year increases, indicating that the model fits the data better for early calendar years. Also, the plot of residuals against the ages shows equivalently bad fit across all of them. The CBD model shows excessive residuals values for the more recent years of the data-set. Additionally, the respective panel of the figure displays a waving pattern when residuals are plotted against the ages of the dataset, with particularly bad fit for certain years, which yield the distinct high values. The plots of the residuals values against the years of birth is dominated by a sort of waving pattern, which emphasises the lack of cohort effects in both models. The waving trend is more intense for the LC model, but also evident for the CBD model. Furthermore, both plots make more evident the observations noted in Chapter 1 about the magnitude of the residuals corresponding to the cohorts born in years 1918, 1919, 1920.

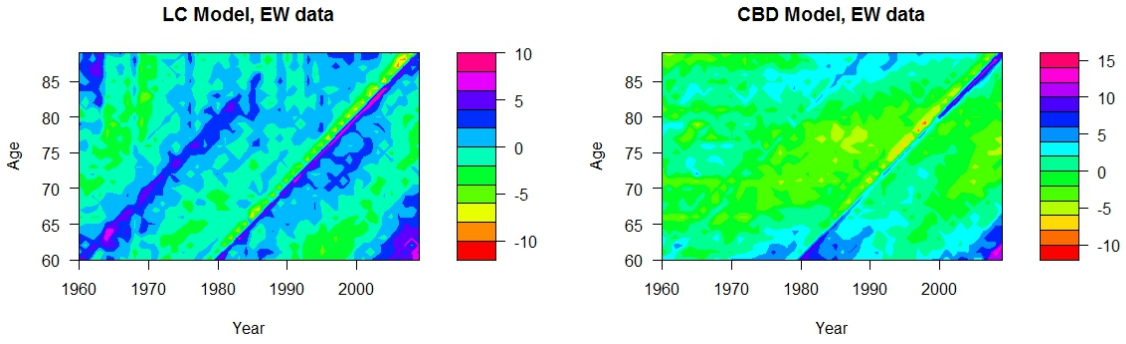

Figure 3.1: Contour plots of standardised residuals of the LC and the CDB models on the left and right, respectively, EW data.

3.2.2 Statistical Analysis

The observed violations of the underlying assumptions are further examined by testing certain hypotheses regarding the series of residuals with respect to the ages and years of the dataset. The matrices are broken down as described in the introduction of this Chapter with the noted statistical tests conducted at the 5% confidence level. Table 3.1 illustrates the rejection rates for each of the employed tests for both models. The results indicate that the independence hypothesis can be rejected at least 6 out of 10 times across both dimensions for both models. Notably, the zero mean hypothesis can never be rejected, except for the age series of the CBD model, for which it was always rejected. In contrast, the unitary variance requirement was never supported by the estimated residuals across any of the two dimensions. The normality tests exhibit significantly lower rejection rates for the CBD model in both the age and year dimensions, but more prominently when considering the years.

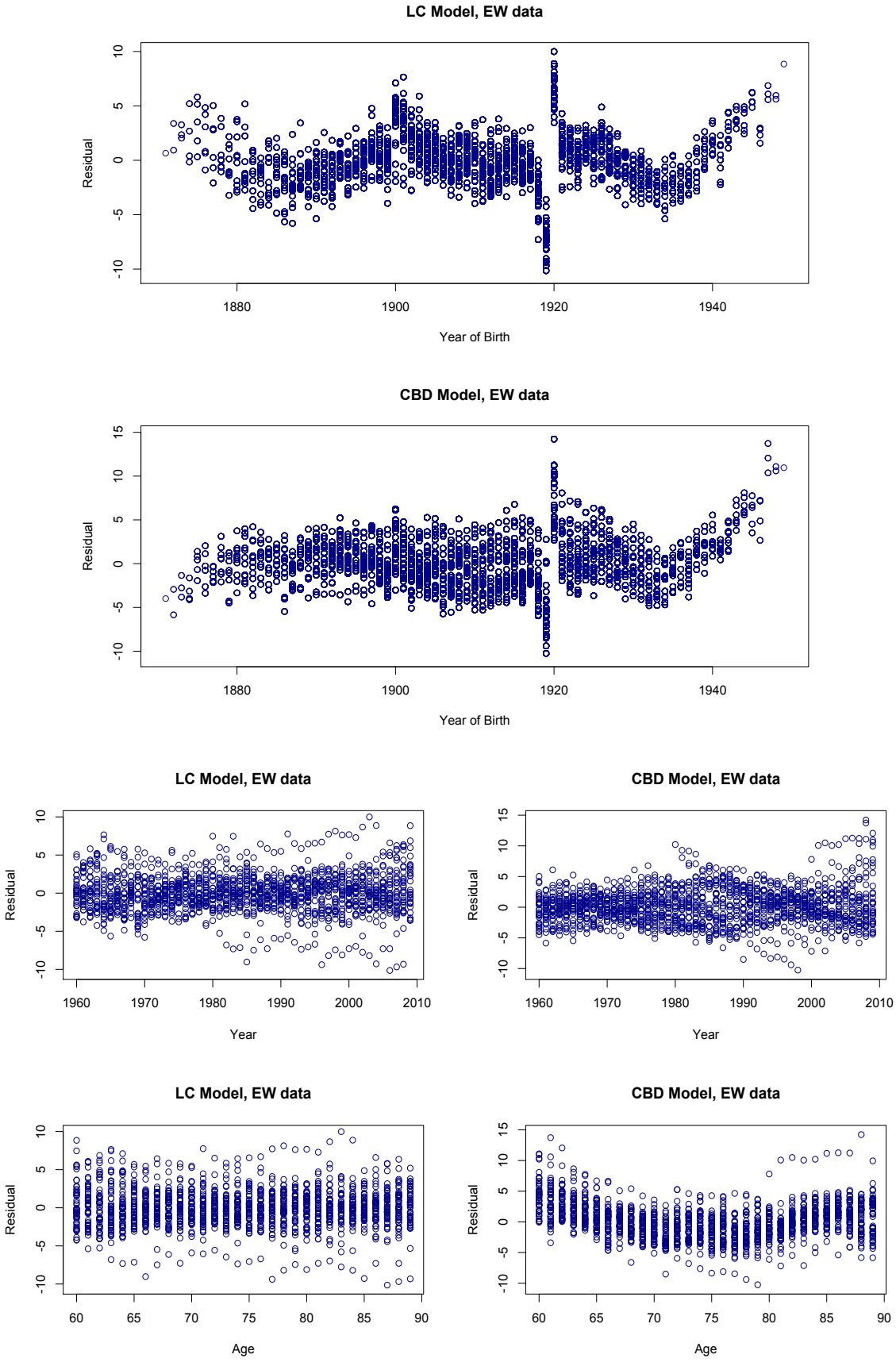

Figure 3.2: Standardised residuals of the LC and CBD models against the years, the ages and the years of births of the cohorts, EW data.

The latter point reveals an important difference between the LC and CBD models. The CBD model has two time-varying factors which capture the mortality dynamics more efficiently than the LC model, whose single period factor is insufficient in that respect. This leads the examined residuals to behave better regarding the normality and independence assumptions for fixed year, except for the case of examining the age series of the CBD model. However, it should be noted that this behaviour is strongly related with the general structure of the (log/logit) death/mortality rates at higher ages; that is, they are strongly linear with age. Nevertheless, the CBD structure is designed specifically to model rates at higher ages and therefore, it is reasonable that it behaves better for the age bands used in this implementation.

	LC Model		CBD Model	
	Age Series	Year Series	Age Series	Year Series
Runs test	60%	66%	63.30%	62%
JB test	66.67%	64.71%	45.45%	26.3%
t-test	0%	0%	100%	0%
χ^2-test	100%	100%	100%	100%

Table 3.1: Rejection rates of the statistical tests at a 5% significance level for the standardised residuals matrices of the LC and CBD models, EW data.

3.2.3 Correlation Analysis

The exact correlation structure between the distinct age and years series is now analysed. Initially, we examine whether the Pearson correlation coefficient of two successive series is significant at the 5% confidence level. Further, we assess the estimated cross-correlation functions between successive residuals series and inspect the lags at which the higher values arise.

For each model there are 78 such Pearson coefficients to examine. 49 are created from the successive years series; that is, the residual vectors across all ages between years 1960 & 1961, 1961 & 1962 and so on. Similarly, there are 29 more tests for the residuals series across all years of the dataset between ages 60 & 61, 61 & 62 and so on. It should be noted that the Pearson coefficient depicts correlation in the case of linear dependence between the variables. Other measures, such as rank based correlation coefficients, would give better insight into the reality, which might be non-linear. However, the residuals will be modelled with a linear model, and thus, the Pearson coefficient seems to be a more appropriate index for our purposes. For

the LC residuals matrix, the coefficient appears statistically important in 42% of the tests between successive age series, and in 70% of the cases for successive years series. The correlation coefficient between successive age series of the CBD model turns out to be significant 45% of the times tested, and the respective rate for years series is 30%. The latter percentage points out once more the beneficial effect of the bivariate random walk driving the dynamics of the CBD model.

The cross-correlation functions of the identical paired vectors as before are also estimated. These yield the amount of correlation between successive series of residuals across various lags. In terms of age series, both models develop the highest value for lag equal to -1. That is, if $R(x,t)$ is the residual for age x and year t, then for all ages x the value of (the) age $x-1$ series to which $R(x,t)$ is more correlated is $R(x-1,t-1)$. This is, of course, exactly due to the cohort effect within the residuals matrices of both models. On the other hand, the results are different between the two models when considering year series. The matrix of the CBD model continues to yield the highest correlation values for lag -1, reconfirming the cohort effect. The matrix of the LC model produces maxima separated between lags 0 and -1, with the former lag indicating the existence of additional age factors in the matrix.

Finally, the autocorrelation functions of the constructed age series are examined with attention to those which reject the independence null hypothesis of the LB tests. The plots of those functions exhibit sinusoidal decaying structure, which tails off the significance level after the first few lags. On the other side, their respective PACF's indicate significant values, if any, at very low lags.

3.3 The Residuals Process

The dependence structures of the examined residuals matrices may possibly be better captured by a single multivariate stochastic process in time rather than by including further age, period, cohort and interaction terms in the basic mortality models.

So, instead of considering families of mortality models of the form

$$\log\big(m(x,t)\big) = \sum_i \beta_x^{(i)} \kappa_t^{(i)} \gamma_c^{(i)}, \tag{3.3.1}$$

for the LC case for example, to improve their in-sample explanatory properties, we prescribe extensions of the form

$$\log\big(m(x,t)\big) = \log\big(m^*(x,t)\big) + \mathcal{R}(x,t),$$

where $m^*(x,t)$ are for this case the LC rates. Similarly, the method may be applied for the CBD family as

$$\text{logit}\big(q(x,t)\big) = \text{logit}\big(q^*(x,t)\big) + \mathcal{R}(x,t).$$

The introduction of the latent states $\mathcal{R}(x,t)$ as components ensures the fitting properties of the resultant model. In such a case, the residuals are considered as an integral part of the stochastic mortality model and are estimated jointly with all other latent states. Modelling the residuals as additional latent states allows for introducing richer interdependencies within the cells of the population than the persistent structure of the cohort effects. For example, if the assumption about the constant contribution of the cohort parameters in the rates across the diagonals of the mortality table was really accurate, one would expect under a model that does not include cohort effects to observe residual series on the same diagonal having low variability, i.e. the constant cohort effect plus the presumed *i.i.d.* noise. However, examination of residual series of the same cohort, for both the residuals of the LC and CBD models, reveals persistently excessive standard deviation values, which could justify the incorporation of some less persistent modelling framework. If the cohort effects within the residuals structure are primarily driven by lifestyle factors, such as smoking, then it is likely that there will be some diffusion between adjacent cohorts, which will be captured by the residuals model. By utilising an appropriate multivariate residuals model one may assume various correlation structures across both dimensions of the mortality table. If the matrix $\mathcal{R}(x,t)$ may be described by a parsimonious stationary stochastic model, the residuals will be predictable and hence the long-term variability of the new model will be driven by the original dynamics of the benchmark model used. Therefore, the non-stationary stochastic factors that mortality depends on are greatly reduced. For example, in the case of the augmented LC family there would be a single such factor. Short-term or local dependencies within the cells of the data-set are reproduced by relating each $\mathcal{R}(x,t)$ term to adjacent cells entries according to a simple linear model. The residuals model may be structured to capture the cohort effect as well as the further dependencies detected in Section 3.2 which are not captured by the currently used age, period and their interaction terms. The proposed framework implies that there are long-term factors, such as the period effects of the CBD model, which develop the main shape of the mortality surface, while there also exist local, short-term dependencies between the cells of a data-set. The short term dynamics carry the disadvantage of data requirement. In particular, for as many ages included in the data-set, we need at least so many years of observation. However, the robustness of a model and the structural changes of mortality over time are often examined by taking a small set of years and contrasting to a larger hyper set. In such an exercise, some

compromise have to be made similarly to that of Section 6.6.

In this section the structure of such a potential model is explored by using residual estimates for both the LC and CBD models from the EW data. The analysis will give a useful insight into the appropriate assumptions that should be made for developing the models of the following Chapters.

3.3.1 Model Description

Let $\mathcal{R}(x,t)$ now denote the latent states which correspond to the estimated raw mortality residuals. For example, for the LC model and if hatted parameters indicate estimated quantities, estimates for the latent states just defined are given as:

$$\mathcal{R}(x,t) = \log\left(\widetilde{m}(x,t)\right) - \widehat{\beta}_x^{(1)} - \widehat{\beta}_x^{(2)}\widehat{\kappa}_t^{(2)}. \tag{3.3.2}$$

Based on the observations and results of Section 3.2 we define a multivariate autoregressive relationship within the elements of the latent residual states as follows:

$$\begin{aligned}
\mathcal{R}(1,t) &= \alpha_1\mathcal{R}(1,t-1) + e_t^{(1)} \\
\mathcal{R}(2,t) &= \alpha_2\mathcal{R}(1,t-1) + \alpha_3\mathcal{R}(2,t-1) + e_t^{(2)} \\
\mathcal{R}(x,t) &= \alpha_4\mathcal{R}(x-2,t-1) + \alpha_5\mathcal{R}(x-1,t-1) + \alpha_6\mathcal{R}(x,t-1) + e_t^{(x)},
\end{aligned} \tag{3.3.3}$$

for $t = 2,\ldots,n$ and $x = 3,\ldots,m$, and where the $e_t^{(1)}, e_t^{(2)}$ and $e_t^{(x)}$ terms are *i.i.d.* error terms through time, but possibly correlated across the ages.

The first of the above equations represents a classical first order autoregressive process for the residual series of the first age in the data-set. Essentially, the residuals of the log-rates for the initial entrants of the population are regressed on the residuals the previous adjacent cohort produces. For the second age in hand, an analogous regression is employed; in this case the residuals are regressed on both those produced by the previous adjacent and the exact cohort the entry belongs to. Therefore, this equation includes the previously applied autoregression, together with the impact of the cohort effect for the second age in hand, measured by α_2. Finally, the last equation gives the general form of the model, according to which each entry of the residual matrix is expressed as a linear combination of the residuals of the previous, the next and the cohort the entry belongs to.

The model defined in equations (3.3.3) can be expressed in vector form, $\boldsymbol{\mathcal{R}}_t$, so that they represent residuals the log-rates for fixed year, across all ages. In such a case,

the model takes the form of a Vector Autoregressive (VAR) process, written as:

$$\boldsymbol{\mathcal{R}}_t = \mathcal{A}\boldsymbol{\mathcal{R}}_{t-1} + \boldsymbol{Z}_t, \tag{3.3.4}$$

for $t = 2, \ldots, n$, where $\boldsymbol{Z}_t \sim N_m\big(0, V_Z\big)$ are *i.i.d.* vectors for all t of dimension m, where m coincides with the number of ages used in implementing the model, and V_Z is the covariance matrix of the disturbances, so that they are independent through time but possibly contemporaneously correlated. The autoregressive matrix, $\mathcal{A}$, takes the form:

$$\mathcal{A} = \begin{pmatrix} \alpha_1 & 0 & 0 & 0 & \ldots & \ldots & \ldots & 0 \\ \alpha_2 & \alpha_3 & 0 & 0 & \ldots & \ldots & \ldots & 0 \\ \alpha_4 & \alpha_5 & \alpha_6 & 0 & \ldots & \ldots & \ldots & 0 \\ 0 & \alpha_4 & \alpha_5 & \alpha_6 & \ldots & \ldots & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ldots & \ddots & \ldots \\ 0 & \ldots & \ldots & 0 & \alpha_4 & \alpha_5 & \alpha_6 & 0 \\ 0 & \ldots & \ldots & 0 & 0 & \alpha_4 & \alpha_5 & \alpha_6 \end{pmatrix}.$$

Under the suggested structure of the autoregressive matrix $\mathcal{A}$, the coefficients α_2 and α_5 are intended to capture the cohort-type effects of the mortality data. Then, if cohort effects are primarily driven by lifestyle factors, such as smoking, then it is likely that there will be some diffusion between adjacent cohorts, which should then be reflected in diminishing values of, for example, the values of α_4 and α_6 in the case of coefficient α_5. Furthermore, the precise specification of the latter matrix, V_Z, relies heavily on the benchmark model and dataset to which the suggested methodology is applied to, as will be clearly shown in later sections.

The objective of the suggested modelling is to specify a stationary model for the defined latent residual states, $\mathcal{R}(x,t)$. That attribute will lead the resulting new model mainly to be driven by the stochastic dynamics of the original model, for each case considered, since the contribution of the additional residuals process will be constant over time.

3.3.2 Empirical Analysis

Before applying the suggested modelling extension to the LC and CBD models, it is useful to conduct some empirical analysis to their estimated raw log/logit mortality residuals, as appropriate. For example, the matrix serving as observations for the LC model is given by equation (3.3.2). This will give a better insight into the exact properties of the suggested model, but also it is used as a guide for particular prior specifications. The parsimonious structure of the autoregressive matrix $\mathcal{A}$ yields

simple least squares estimates for the autoregressive parameters. For example, the estimator for α_1 is given as:

$$\widehat{\alpha}_1 = \frac{\sum_{i=1}^{n-1} \mathcal{R}(1,i+1)\mathcal{R}(1,i)}{\sum_{i=1}^{n-1} \mathcal{R}^2(1,i)}. \tag{3.3.5}$$

Further, let the $(n-1)\times 1$ vector $\mathbf{Y}_1 = \mathcal{R}(2,t_y)$, for $t_y = 2,\ldots,n$, the $(n-1)\times 2$ matrix $\mathbf{W}_1 = \big(\mathcal{R}(1,t_w), \mathcal{R}(2,t_w)\big)$, for $t_w = 1,\ldots,n-1$, and the 2×1 vector $\mathbf{a} = (\alpha_2,\alpha_3)^{'}$. The estimator for α_2 and α_3 are given as:

$$\begin{pmatrix} \widehat{\alpha}_2 \\ \widehat{\alpha}_3 \end{pmatrix} = \big(\mathbf{W}_1^{'}\mathbf{W}_1\big)^{-1}\mathbf{W}_1^{'}\mathbf{Y}_1. \tag{3.3.6}$$

Finally, the least squares estimates for the rest parameters of the matrix $\mathcal{A}$ are derived by exploiting appropriate blocks of the raw residuals matrix. Define the following blocks:

$$\begin{aligned} \mathcal{R}_1 &= \mathcal{R}(x_1,t_y),\ \ \mathcal{R}_2 = \mathcal{R}(x_1,t_w) \\ \mathcal{R}_3 &= \mathcal{R}(x_2,t_w),\ \ \mathcal{R}_4 = \mathcal{R}(x_3,t_w), \end{aligned}$$

where $x_1 = 3,\ldots,m$, $x_2 = 2,\ldots,m-1$, $x_3 = 1,\ldots,m-2$ and t_y and t_w as before. Further, let $\mathbf{Y}_2 = vec\big(\mathcal{R}_1\big)$, $\mathbf{W}_2 = \Big(vec\big(\mathcal{R}_2\big), vec\big(\mathcal{R}_3\big), vec\big(\mathcal{R}_4\big)\Big)$ and $\mathbf{b} = (\alpha_4,\alpha_5,\alpha_6)^{'}$, where *vec* is the operator which stacks the columns of a matrix in a single vector one by one. The required estimates are given in accordance to those of equation (3.3.6), as:

$$\begin{pmatrix} \widehat{\alpha}_4 \\ \widehat{\alpha}_5 \\ \widehat{\alpha}_6 \end{pmatrix} = \big(\mathbf{W}_2^{'}\mathbf{W}_2\big)^{-1}\mathbf{W}_2^{'}\mathbf{Y}_2. \tag{3.3.7}$$

The estimators from equations (3.3.5), (3.3.6) and (3.3.7) yield the least squares solution of $\mathcal{A}$, $\widehat{\mathcal{A}}$, and based on multivariate time series theory (Lütkepohl, 2005), they are asymptotically equivalent to the MLE's. The estimates from the applied regression to the appropriate matrices of the LC and the CBD models are shown in Table 3.2 overleaf. The estimate $\widehat{\alpha}_1$ is reasonably high since there is a single regressor for the series. The estimate for the vector $\mathbf{a}$ shares common characteristics among the two models, but also develops distinct behaviour between them. Firstly, estimate $\widehat{\alpha}_2$ is the largest entry of the vector for both models, which is intuitive since it concerns the cohort-type effect within the regression for the second age in the dataset, which is known to be the main driver of the assigned dynamics. However, in the case of the LC model, the estimate $\widehat{\alpha}_3$ is firstly negative, but most importantly negligible in absolute value compared to $\widehat{\alpha}_2$. For the CBD model on the other hand, the estimate

$\widehat{\alpha}_3$ is about half of $\widehat{\alpha}_2$, which might imply that the model fails to capture existing age effects within the EW data-set. In turn, this might be due to the simple functional age effects it includes. In retrospect, this may also be the reason for the difference in the magnitude of the $\widehat{\alpha}_1$ estimate between the two models.

	LC Model	CBD Model
$\widehat{\alpha}_1$	0.7351	0.9759
$\widehat{\alpha}_2$	0.8715	0.6121
$\widehat{\alpha}_3$	-0.00597	0.3425
$\widehat{\alpha}_4$	0.0189	0.0537
$\widehat{\alpha}_5$	0.6921	0.6627
$\widehat{\alpha}_6$	0.0202	0.1985

Table 3.2: Least squares estimate of the autoregressive matrix, $\mathcal{A}$, from empirical residuals matrices of the LC and CBD models, EW data.

Finally, the estimate of the vector $\mathbf{b}$ gives the coefficients of the autoregression for all the ages in the data-set, except for the first two. The estimate $\widehat{\alpha}_5$ is the largest among the entries of the vector consistently, since it is the cohort-type factor. The impact of the previous year born adjacent cohort, through $\widehat{\alpha}_4$, is small, but much larger for the CBD model. In accordance to the other two autoregressive relationships examined, the estimate of $\widehat{\alpha}_6$ is about 30% of $\widehat{\alpha}_5$ for the CBD model, while only 3% of $\widehat{\alpha}_5$ in the case of the LC model.

The empirical estimate of the autoregressive matrix $\mathcal{A}$ aids in identifying the properties of the residuals model. In particular, the VAR process is called stable and thus stationary if and only if the eigenvalues of $\mathcal{A}$ are in the interval (-1,1) (Lütkepohl, 2005). The requirement is satisfied for the empirical matrices of the LC and the CBD models. The assumptions regarding the disturbances, $\boldsymbol{Z}_t$, are tested in retrospect by their estimated residuals, given as:

$$\widehat{\mathbf{Z}}_{t+1} = \boldsymbol{\mathcal{R}}_{t+1} - \widehat{\mathcal{A}}\boldsymbol{\mathcal{R}}_t, \tag{3.3.8}$$

for $t = 1, \ldots, n-1$. The multivariate version of the Shapiro-Wilks statistic (Villasenor Alva and Estrada, 2009) is employed to test the Normality assumption about the series of estimates $\widehat{\mathbf{Z}}_t$. For both matrices the p-value of the empirical sample is nowhere close to any importantly low statistical significance level. Further, the zero mean hypothesis about the Normal distribution of the disturbances is assessed with the T^2-Hotelling statistic (Hotelling, 1992), which in both cases yields p-values almost equal to one.

Summing up, the empirical analysis of the raw residuals of the LC and CBD models from their application to the EW dataset, shows that their essential properties may be modelled with a simple stationary multivariate autoregression. Since the expectation of the disturbances, $\mathbf{Z}_t$, is the zero vector, the same is true for the time invariant mean of the model. Further, the stationary autoregressive structure of the model yields a diminishing, independent of time covariance matrix for the residuals process, $V_{\mathcal{R}}$. Its specification solely depends on the autoregressive matrix $\mathcal{A}$ and the one-step ahead covariance of the disturbances, V_Z, and it is taken as the solution of the system:

$$V_{\mathcal{R}} = \mathcal{A} V_{\mathcal{R}} \mathcal{A}^{'} + V_Z, \tag{3.3.9}$$

which is linear in the elements of $V_{\mathcal{R}}$, but of dimension $m^2 \times m^2$, where m is the common dimension of all the matrices involved.

The maximum likelihood estimate of the matrix V_Z is taken from the estimated residuals of equation (3.3.8) as:

$$\widehat{V}_Z = \frac{1}{n-1} \sum_{i=1}^{n-1} \widehat{\mathbf{Z}}_{t+1} \widehat{\mathbf{Z}}_{t+1}^{\prime}. \tag{3.3.10}$$

The form of the covariance matrix V_Z highlights the structural differences between the LC and CBD models. Figure 3.3 shows the scaled empirical estimates of matrix $\widehat{V}_Z$, so that they represent the underlying correlation. The left panel develops positive residuals correlation for neighbouring, or at least not very distant, ages at both sides of the data-set. At around the middle of the ages of the data, the residuals correlation becomes negative and decreases further, so that residuals of ages at the start are strongly negatively correlated to corresponding values of ages at the end of the data. On the other hand, in the case of the CBD model in the right panel, there are fluctuating positive and negative correlation values with no identifiable structure.

The range of the values is evidently lower than in the case of the LC model, and the main reason behind their distinct behaviour is the two stochastic factors of the CBD model. The bivariate random walk driving the logit mortality rates under the CBD model is able to capture simultaneously diverse dynamics and thus, applying the suggested VAR model for its residuals, yields undeterminable and lower estimated correlation structure for the disturbances. The single factor of the LC model is structurally incapable of capturing that relationship, which becomes apparent when the covariance matrix of the disturbances of the residuals process is estimated.

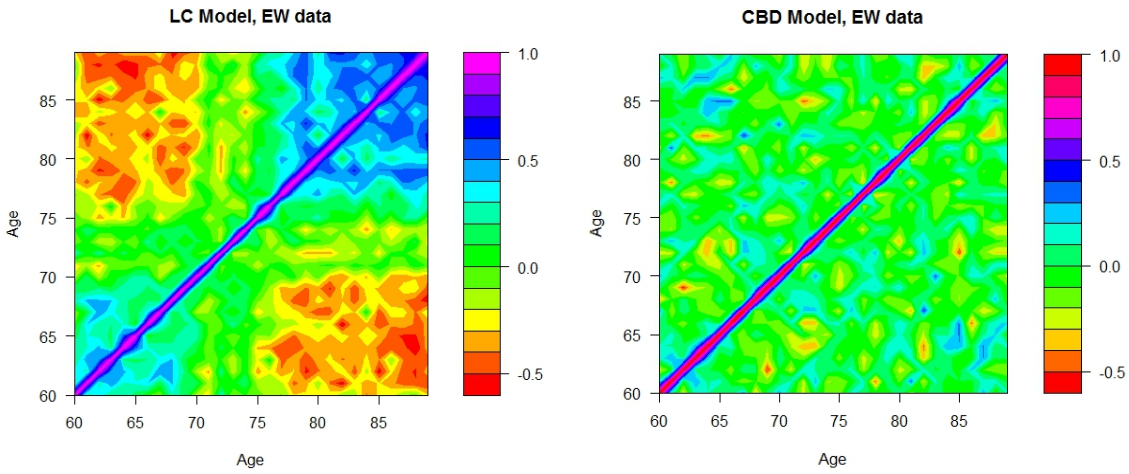

Figure 3.3: Empirical one-step ahead correlation matrices of the applied autoregression for the raw residuals of the LC and CBD models, EW data.

Another difference in the residuals model between the two benchmark models may be seen if the variance components of the empirical covariance matrix, $\widehat{V}_Z$, are plotted. Figure 3.4 shows the diagonal elements of matrix $\widehat{V}_Z$ under the LC and CBD models in black and red, respectively. There are two points to note in the graph. First, the high variability for the initial ages of the data-set under both mortality modelling structures. This is due to a combination of reasons. The Poisson distribution assumes that deaths have equal mean and variance, but this is hardly ever true for young, and also for old, ages where deaths are rare and the variability high. Moreover, the structure of the residuals model implies one and two regressors for the first and second ages of the data-set, respectively. The high variance for the first age under both models is most likely due to the single regressor. Second, the distinctive behaviour of the residuals of the LC model for old ages of the population, in contrast to the behaviour of the residuals of the CBD model. The main reason for that difference is the two-factor structure of the CBD model. The second stochastic factor of the CBD model is capable of capturing the variability of deaths at high ages; hence, the variance components of the latent residuals are do not need to compensate as they do under the single factor structure of the LC model.

3.4 Summary

This Chapter developed the fundamental ideas underlying the modelling framework suggested by the Thesis. By examining the statistical properties and correlation structures within the estimated mortality residuals of the LC and CBD models for the EW data, it is argued that the information included in the residuals may be described by a multivariate stochastic process. The assessment of the estimated mortality residuals

aided in specifying a parsimonious stationary VAR model, which might be used to augment any of the two examined structures. The model is capable of capturing efficiently the cohort effect mortality data usually include, while introducing short-term dependencies between the cells of the population. As will be shown in Chapter 6, the accommodation of the suggested augmentation within a simple stochastic model is also able of carrying forward the full mortality characteristics of the modelled population in the projections.

The proposed bespoke regression behaves differently under the two distinct benchmarks. The empirical analysis of the estimated residuals shows that although the interpretation of the autoregressive coefficients is kept consistent, the associated covariance structure varies. The two stochastic factors of the CBD model allow for an easily describable, maybe diagonal form for the matrix V_Z. On the other hand, the LC benchmark develops a specific pattern which would need to be estimated within the algorithm. That attribute of the residuals model for the mortality residuals of the CBD model results in a simpler model specification and associated estimation algorithm in Chapter 5. Finally, the statistical analysis of the empirical results verifies the distributional assumptions underlying the residuals model.

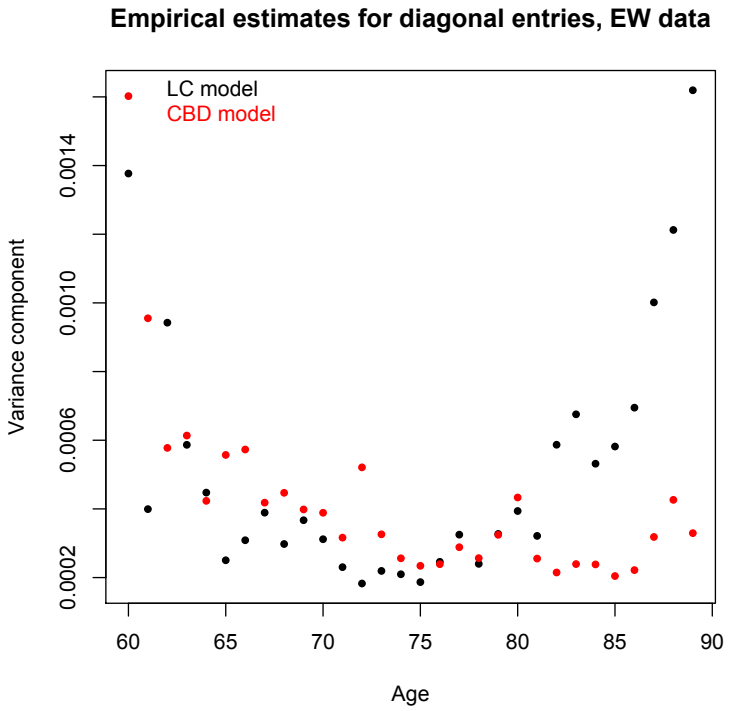

Figure 3.4: Diagonal elements of empirical covariance matrix, $\widehat{V}_Z$, of the VAR residuals model under the LC and CBD benchmarks, EW data.

Chapter 4

Bayesian Augmented Lee-Carter Model

By assuming that mortality patterns may be explained by effects of different level of persistence, this Chapter accommodates local dynamics under the single-factor structure of the LC model. On the one hand, the persistent stochastic factor, $\kappa_t^{(2)}$, and the fixed age effects, $\beta_x^{(1)}$ and $\beta_x^{(2)}$, determine the main shape of the mortality surface. On the other hand, the short-term dependencies between the cells of a data-set are modelled in a new way, by augmenting the modelling structure with an additional multivariate stochastic process for the residuals. The stationary residuals model of Chapter 3 might incorporate local interdependencies within the data, while it is also able to capture efficiently the cohort effect. The long-term dynamics of the delivered model are still governed by the stochastic period factor $\kappa_t^{(2)}$. Under the suggested augmentation and the Bayesian estimation approach adopted, the method is named Bayesian Augmentation (BA) and the developed model is termed BA-LC. The resulting hierarchical structure of the model is illustrated and the related estimation procedure is presented. The model is applied to the EW and Canada data. The convergence of the associated algorithms is assessed and the posterior distributions of the BA-LC model is summarised. The results are compared to those of the original LC and APC models under the conditionally independent Poisson model for deaths from Chapter 1.

4.1 Model description

The regression relationship of the LC model in each cell is supplemented with one additional residual state, $\mathcal{R}(x,t)$. The central death log-rates for age x last birthday, during calendar year t are given as:

$$\log\big(m(x,t)\big) = \beta_x^{(1)} + \beta_x^{(2)}\kappa_t^{(2)} + \mathcal{R}(x,t). \tag{4.1.1}$$

Through the introduction of the latent residual terms, each log-rate is affected by an additive factor $\mathcal{R}(x,t)$ relative to the original LC model. The latent residuals, $\mathcal{R}(x,t)$, are estimated jointly with all the other parameters of the model, and are presumed to be generated by an additional stochastic model, which acts independently of the other stochastic factors of the model, namely $\kappa_t^{(2)}$ in this case. The Poisson assumption for the deaths, $D(x,t)$, is kept, so that it represents the main source of idiosyncratic risk in the model. However, we depart from the conditionally independent Poisson assumption underpinning the models of Chapter 1, and model the latent residual states as dependent variables under the stochastic process of Chapter 3. The common choice of the Poisson model will facilitate comparison between the developed models with the models of Chapter 1, in Chapter 6. The error structure imposed by the VAR model described in Chapter 3 introduces serial correlation within the cells of the modelled mortality data which in the mortality context is interpreted as short-term dependencies within the population. Due to the autoregressive nature of the residuals model, its affect in the forecasted rates is constant and small relative to the other parts that construct the future rates as the projection horizon extends. Specifically, each vector $\boldsymbol{\mathcal{R}}_t$ will be contributing a sample from a $N_m\big(0, V_{\mathcal{R}}\big)$ distribution, as discussed in Section 3.3.2. The long term dynamics of the model are still governed by the random walk model for the period factor $\kappa_t^{(2)}$, so that:

$$\kappa_{t+1}^{(2)} = \kappa_t^{(2)} + \delta + \zeta_{t+1},$$

where the disturbances $\zeta_t \sim N(0, \sigma_\kappa^2)\ \forall t$, and independent of any other random components of the model.

The hierarchical structure of the full model is developed as follows. On the top level, the deaths follow the Poisson assumption as:

$$D(x,t)|m(x,t) \sim Poi\big(E(x,t)m(x,t)\big). \tag{4.1.2}$$

The log-rates are connected to the latent parameters of the model through the link equation (4.1.1), so that the likelihood might be derived. Furthermore, the distribu-

tional assumptions for the stochastic components of the model are as follows:

$$\kappa_{t+1}^{(2)}|\kappa_t^{(2)}, \delta, \sigma_\kappa^2 \sim N\big(\kappa_t^{(2)} + \delta, \sigma_\kappa^2\big) \tag{4.1.3}$$

$$\boldsymbol{\mathcal{R}}_{t+1}|\boldsymbol{\mathcal{R}}_t, \mathcal{A}, V_Z \sim N_m\big(\mathcal{A}\boldsymbol{\mathcal{R}}_t, V_Z\big) \tag{4.1.4}$$

$$\boldsymbol{\mathcal{R}}_1|\mathcal{A}, V_Z \sim N_m\big(0, V_{\mathcal{R}}\big). \tag{4.1.5}$$

Equation (4.1.3) refers to the conditional normality of the random walk model for the period effects, $\kappa_t^{(2)}$. Equations (4.1.4) and (4.1.5) are related to the multivariate residuals process. The former corresponds to the multivariate conditionally normal distribution of the residuals vectors, $\boldsymbol{\mathcal{R}}_t$, for all years t, except the first. The latter relates to the first observation year of the dataset. The vector $\boldsymbol{\mathcal{R}}_1$ is assigned the stationary distribution of the multivariate autoregressive model, under the assumption that it exists. The above specification and the validation of the corresponding assumption is ensured by the solution of the system which yields the matrix $V_{\mathcal{R}}$, given in equation (3.3.9), at each iteration of the stochastic simulation employed to estimate the model. If this is the case, $\boldsymbol{\mathcal{R}}_t$ is indeed stationary and in the long-term contributes random vectors with expectation the zero vector and constant covariance matrix, determined by the convergent solution of the linear system for $V_{\mathcal{R}}$ in equation (3.3.9).

4.2 Model fitting

In this Section the Bayesian implementation of the augmented LC model is presented. Due to the augmentation of the mortality modelling structure with the additional latent residuals $\mathcal{R}(x,t)$, and the Bayesian implementation employed, the resulting model is referred to as Bayesian Augmented LC (BA-LC) model. Similarly, Chapter 5 develops the BA-CBD model, and the BA-models are compared in Chapter 6. The MCMC algorithm used to estimate the BA-LC model is developed, and the required components for its implementation are derived.

The estimation of the model follows from the accompanying stochastic assumptions, given in equations (4.1.2)-(4.1.5). The global parameter vector of the model, $\boldsymbol{\theta}$, comprises the latent age, period and residual effects, along with the parameters for each of the models assigned for $\kappa_t^{(2)}$ and $\boldsymbol{\mathcal{R}}_t$.
Therefore, let

$$\boldsymbol{\theta} = \Big\{\beta_x^{(1)}, \beta_x^{(2)}, \kappa_t^{(2)}, \boldsymbol{\mathcal{R}}_t, \delta, \sigma_\kappa^2, \mathcal{A}, V_Z\Big\}.$$

The designed model has more unknowns than data points. Essentially, each data point is expressed in terms of the latent states, $\beta_x^{(1)}, \beta_x^{(2)}, \kappa_t^{(2)}$ and $\boldsymbol{\mathcal{R}}_t$, where the $\boldsymbol{\mathcal{R}}_t$

states alone are as many as the data points. Furthermore, the parameter space of the model includes the covariance matrix V_Z, which is not particularly structured. As shown in the end of Chapter 3, the diagonal variance elements should have a U-shaped pattern across the examined ages to compensate for the lack of a second stochastic factor in the LC model. Furthermore, V_Z needs to be developing a decaying correlation structure where positively signed relationship at young ages turns to negative as the difference between the examined ages increases. Inventing a parameterisation which combines the above characteristics, along with being stable enough in terms of the required calculations, as for example the system of equation (3.3.9), is challenging and definitely needs to be bespoke for each examined case/data-set. The hierarchical structure of the proposed latent states model advocates the Bayesian paradigm. The Bayesian approach allows for joint estimation of the latent states and parameters of the underlying stochastic models, hence reducing modelling to a single stage procedure. Prior assumptions that are consistent with the specifications of the developed model aid in deducing satisfying approximations for the posterior distributions of its components. MCMC methods are used to sample from the full posterior distribution of the model, and the obtained sample allows for comprehensive inference and coherent incorporation of parameter uncertainty in the relevant mortality projections.

The Bayesian estimation of the structure developed to determine the dynamics driving mortality assigns the correct degree of associated uncertainty for all the quantities involved in the model. For example, the age effects, $\beta_x^{(1)}$ and $\beta_x^{(2)}$, are now inherently subject to some degree of variation, expressed through their simulated posterior distributions. The same stands for all other factors included in the model and the corresponding pre-specified stochastic models' parameters. Those distributions are used to incorporate additional risk in the projected rates. This is in contrast to the estimation and projection method employed in Chapter 1, where the latent states and the parameters of the stochastic models were assumed to be fixed unknowns. The suggested method of incorporating parameter uncertainty is closer to a full Bayesian solution, such as those developed in Czado et al. (2005) and Pedroza (2006) and the Bayesian version of the CBD model developed in Section 2.4.1. That is, the problem is being set in a unified Bayesian framework and a holistic estimation is applied according to some pre-specified assumptions about the underlying processes of the model.

4.2.1 Implementation

As developed in Chapter 2, the Bayesian paradigm consists of combining the likelihood of the model with prior specifications for its parameters, so that the full posterior distribution is derived. The likelihood follows from the assumptions stated in equations

(4.1.2)-(4.1.5). The log-likelihood is of the form:

$$\ell(\boldsymbol{\theta}) = \ell_1(\boldsymbol{\theta}) + \ell_2(\boldsymbol{\theta}) + \ell_3(\boldsymbol{\theta}) + C,$$

where C is some constant independent of the components of the parameter vector $\boldsymbol{\theta}$.

For convenience and to emphasise the dependence on the parameter vector $\boldsymbol{\theta}$, let $u(\boldsymbol{\theta})$ denote the modelling equation in terms of the latent states of the model, so that:

$$u(\boldsymbol{\theta}) \equiv u(\boldsymbol{\theta}; x, t) = \log\big(m(x,t)\big) = \beta_x^{(1)} + \beta_x^{(2)}\kappa_t^{(2)} + \mathcal{R}(x,t).$$

The three components of the log–likelihood are then given by the following equations:

$$\begin{aligned}
\ell_1(\boldsymbol{\theta}) &= \sum_{x,t}\Big\{D(x,t)u(\boldsymbol{\theta}) - E(x,t)\exp\big(u(\boldsymbol{\theta})\big)\Big\},\\
\ell_2(\boldsymbol{\theta}) &= -\frac{1}{2}\Bigg\{\big(n-1\big)\log\big(|V_Z|\big) + \log\big(|V_{\mathcal{R}}|\big) + \boldsymbol{\mathcal{R}}_1^{'}V_{\mathcal{R}}^{-1}\boldsymbol{\mathcal{R}}_1 +\\
&\qquad \sum_{j=1}^{n-1}\Big[\big(\boldsymbol{\mathcal{R}}_{j+1} - \mathcal{A}\boldsymbol{\mathcal{R}}_j\big)^{'}V_Z^{-1}\big(\boldsymbol{\mathcal{R}}_{j+1} - \mathcal{A}\boldsymbol{\mathcal{R}}_j\big)\Big]\Bigg\},\\
\ell_3(\boldsymbol{\theta}) &= -\frac{1}{2}\Bigg\{(n-1)\log\big(\sigma_\kappa^2\big) + \frac{1}{\sigma_\kappa^2}\sum_{j=1}^{n-1}\Big\{\big(\kappa_{j+1} - \kappa_j - \delta\big)^2\Big\}\Bigg\}.
\end{aligned}$$

Although we use the same argument in all of the above parts of the likelihood, it is implied that in each term only the required components of $\boldsymbol{\theta}$ are used.

ℓ_1 comes from the Poisson assumption about deaths in terms of the parameterisation of the mortality rates. ℓ_2 is the likelihood of the VAR residuals model, including both the conditional and unconditional parts of it. ℓ_3 corresponds to the conditional normal likelihood of the random walk model for the period effects, $\kappa_t^{(2)}$.

The same comments apply as in the discussion of the original LC model, so that the sets of parameters $\beta_x^{(2)}$ and $\kappa_t^{(2)}$ need to be constrained. The summation of $\beta_x^{(2)}$'s is still kept equal to 1, while instead of forcing the $\kappa_t^{(2)}$'s to sum up to zero, $\kappa_1^{(2)}$ is fixed to be zero. Thus, the set of constraints used to facilitate convergence of the algorithm are:

$$\sum_x \beta_x^{(2)} = 1 \ \text{ and } \ \kappa_1^{(2)} = 0. \tag{4.2.1}$$

Finally, the last step for the full specification of the Bayesian model is the relevant prior assignments. According to the hierarchical structure developed, prior distributions are assigned to the lowest level free parameters over all layers of the hierarchy.

That is, a joint prior distribution is assigned for the components of the parameter sub-vector

$$\boldsymbol{\phi} = \left\{ \beta_x^{(1)}, \beta_x^{(2)}, \delta, \sigma_\kappa^2, \mathcal{A}, V_Z \right\},$$

whose components are assumed to be mutually independent.

The intention is to combine appropriate prior assignments with the modelling assumptions to allow for the simulation of the posterior distribution. Where it is possible to use limited information in order to allow the data reveal the information they carry, it is done so. For example, $\beta_x^{(1)}$ and $\beta_x^{(2)}$ are assigned uniform priors over the real line, which are uninformative.The $\beta_x^{(1)}$ are conjugate within the model as will be seen further on, while the interaction of $\beta_x^{(2)}$ with $\kappa_t^{(2)}$ yield no distinguishable form of distribution. The parameters of the random walk model, δ and σ_κ^2, are given their conjugate prior pair, which consists of the Normal-IG specification. Even though this is not an non-informative specification, vague information might be supplied through the parameters of the assigned distributions. The advantage of the assignment consists in the posterior tractability of the model which leads in an efficient Gibbs sampling scheme.

The autoregression matrix $\mathcal{A}$ contains six parameters. The first of them, α_1, is assigned a $Beta(2,2)$ prior mapped to the interval $(-1,1)$. The particular specification maps a symmetric parabolic distribution with mean 0.5, from (0,1) to (-1,1) (Johnson et al., 1995). Thus, the autoregressive parameter α_1 is *a priori* centred around the mean 0, and is assigned zero prior probability outside the interval $(-1,1)$, since otherwise the residuals process would not be stationary. Parameters α_3 and α_6 are given Uniform prior distributions on $(-1,1)$ since they should also be restricted in $(-1,1)$. The rest of the parameters within the matrix $\mathcal{A}$ are assigned uninformative uniform improper priors over the real line. Jointly viewed, the prior distribution for matrix $\mathcal{A}$ is only constrained by the zero probability assignment outside interval (-1,1) for α_1, α_3 and α_6. No specific information are supplied, except for the symmetric prior concentration of α_1 around zero. The contradiction of the prior information to the empirical results of Chapter 3 for coefficient α_1 will stress how the data outperform prior assumptions within the Bayesian estimation framework and the implemented method.

Lastly, for the covariance matrix V_Z the Inverse-Wishart (IW) distribution, which is the multivariate analogue of the IG distribution, is used. In accordance with the univariate case, it is the conjugate pair for the covariance matrix of the multivariate normal model, as this arises due to the specification of the residual process. The distribution is determined by two parameters; the degrees of freedom and the positive definite scale matrix. The scale matrix controls its location, while the degrees of

freedom determines the dispersion of the distribution around its location. For some prior assignment such as:

$$V_Z \sim IW_m\left(n_0, \Psi^{-1}\right),$$

where the dimension of both matrices is $m \times m$, the prior mean is:

$$\mathbb{E}(V_Z) = \frac{\Psi}{n_0 - m - 1}.$$

Parameter n_0 needs to be at least $m + 2$ for the distribution to be properly defined. Due to the distribution's sampling construction, values close to $m + 2$ make the distribution loosely centred around its mean, whereas large values for n_0 place the samples from the IW distribution closer to its mean. More information on the IW distribution may be found in classic textbooks regarding multivariate statistics, such as Anderson (1984). The above discussion about the joint prior distribution of the model is summarised in the following assignments:

$$\begin{aligned}
\delta &\sim N(\delta_0, \sigma_\delta^2), \quad \sigma_\kappa^2 \sim IG(a, b), \\
\alpha_1^* &\sim Beta(2, 2) \quad \text{and} \quad \alpha_1 = 2\alpha_1^* - 1, \\
\alpha_3, \alpha_6 &\sim U(-1, 1), \quad \alpha_2, \alpha_4, \alpha_5 \sim U(-\infty, +\infty), \\
V_Z &\sim IW\left(n_0, \Psi^{-1}\right).
\end{aligned}$$

The prior parameters for the drift are set to be $\delta_0 = 0, \sigma_\delta^2 = 1$, so that the prior for the drift is the standard normal distribution. This choice is close to the expectation for the drift, given the experience gained form the LC modelling results. However, it is not strongly informative, but *a priori* locates the drift close to reasonable values. Along similar lines, the vague prior for the variance, σ_κ^2, of the random walk occurs for small values of the parameters of the IG distribution. So, in the implementation of the model these are set to $a = b = 0.01$. Relative to the prior assignment of the covariance matrix V_Z, the matrix Ψ is given a diagonal form with common entry equal to 10^{-4} and degrees of freedom exactly $m + 2$, so that *a priori* the distribution is loosely centred around the specified scale matrix, Ψ. Given the results of Chapter 3, the chosen prior specification is regarded quite uninformative, since it is allowed to vary significantly around its mean, which is not close to the maximum likelihood estimate of V_Z from the empirical residuals analysis. Thus, the data are expected to determine the form of the posterior distribution of the covariance matrix V_Z.

According to the above discussion, the joint prior distribution for the parameter subvector $\boldsymbol{\phi}$ has log-density of the form:

$$\begin{aligned} p(\boldsymbol{\phi}) &= -\frac{1}{2}\Big((n_0+m+1)\log\big(|V_Z|\big) + tr\Big(\Psi V_Z^{-1}\Big)\Big) + \log\big(1-\alpha_1^2\big) \\ &\quad -(a+1)\log\big(\sigma_\kappa^2\big) - \frac{b}{\sigma_\kappa^2} - \frac{1}{2\sigma_\delta^2}\big(\delta-\delta_0\big)^2 + C_2, \end{aligned}$$

for $|\alpha_1|<1$, $|\alpha_3|<1$, $|\alpha_6|<1$, Ψ positive definite and C_2 some normalising constant that does not depend on any of the components of the vector $\boldsymbol{\phi}$, and $tr(\cdot)$ denotes the usual trace of a matrix.

4.2.2 Posterior simulation

The joint log-posterior density of the latent states and the parameters of the BA-LC model is expressed as:

$$\pi(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}) + p(\boldsymbol{\phi}) + C_3,$$

for some constant C_3 independent of the parameter vector $\boldsymbol{\theta}$.

Working with the full conditional log-posteriors reveals several conjugate relationships within the developed structure. For those cases, the Gibbs sampler is employed, whereas for instances where no analytical result arises, the Metropolis–Hastings algorithm, or variants of it, is implemented. The relevant full conditional log-posteriors are now considered in turn for each component, or group of components, of the parameter vector $\boldsymbol{\theta}$. The conditioning in each case is on the global parameter vector except for the parameter, or group of parameters, under consideration, denoted as $\boldsymbol{\theta}_{-i}$, where i is the parameter that is updated, while all other entries are fixed at their latest values.

The full conditional densities of the period effects $\beta_x^{(1)}$ are tractable and for fixed age last birthday, x, we have:

$$\exp\big(\beta_x^{(1)}\big)\,|\boldsymbol{\theta}_{-i} \sim Gamma\left(\sum_t D(x,t), \sum_t E(x,t)\exp\Big(\beta_x^{(2)}\kappa_t^{(2)} + \mathcal{R}(x,t)\Big)\right). \tag{4.2.2}$$

The age-period interaction terms in the log-posteriors of the second set of age effects, $\beta_x^{(2)}$, yield no recognisable form of distribution. Instead, the densities for fixed age last birthday x, are written as:

$$\pi\Big(\beta_x^{(2)}|\boldsymbol{\theta}_{-i}\Big) = \beta_x^{(2)}\sum_t \kappa_t^{(2)} D(x,t) - \exp\big(\beta_x^{(1)}\big)\sum_t E(x,t)\exp\Big(\beta_x^{(2)}\kappa_t^{(2)} + \mathcal{R}(x,t)\Big). \tag{4.2.3}$$

Sampling for those variables within the stochastic simulation is accomplished via normal random walk Metropolis components, centred at the previous state of the chain and appropriately tuned proposal variance.

Along similar lines, the full conditional log-posteriors for the time-series terms of the model, $\kappa_t^{(2)}$ and $\boldsymbol{\mathcal{R}}_t$, are intractable. For both sets of parameters they are mixtures of Poisson and normal components, for fixed calendar year t. However, they take distinct forms for years at the bounds of the dataset, compared to the intermediate years. In particular, all the full conditional log-posteriors of the period effects, $\kappa_t^{(2)}$, involve the corresponding Poisson part of the likelihood, along with normal density terms, derived from the random walk model for the $\kappa_t^{(2)}$ series. The log-densities of $\kappa_1^{(2)}$ and $\kappa_n^{(2)}$ include one such term from the conditional normal density. The log-densities of the intermediate years' period effects include two terms from the normal density beyond their Poisson component.

Sample values of $\kappa_t^{(2)}$'s are achieved as for the age effects $\beta_x^{(2)}$, by normal random walk Metropolis steps, centred at the previous state of the chain and with appropriately tuned proposal variance. The form of the full conditional log-posteriors of all the period effects $\kappa_t^{(2)}$ are as follows:

$$\begin{aligned}
\pi\left(\kappa_1^{(2)}|\boldsymbol{\theta}_{-i}\right) &= \kappa_1^{(2)}\sum_x D(x,1)\beta_x^{(2)} - \sum_x E(x,1)\exp\left(\beta_x^{(1)} + \beta_x^{(2)}\kappa_1^{(2)} + \mathcal{R}(x,1)\right) \\
&\quad -\frac{1}{2\sigma_\kappa^2}\left(\kappa_2^{(2)} - \kappa_1^{(2)} - \delta\right)^2, \\
\pi\left(\kappa_j^{(2)}|\boldsymbol{\theta}_{-i}\right) &= \kappa_j^{(2)}\sum_x D(x,j)\beta_x^{(2)} - \sum_x E(x,j)\exp\left(\beta_x^{(1)} + \beta_x^{(2)}\kappa_j^{(2)} + \mathcal{R}(x,j)\right) \\
&\quad -\frac{1}{2\sigma_\kappa^2}\left[\left(\kappa_{j+1}^{(2)} - \kappa_j^{(2)} - \delta\right)^2 + \left(\kappa_j^{(2)} - \kappa_{j-1}^{(2)} - \delta\right)^2\right], \text{with } 1 < j < n, \\
\pi\left(\kappa_n^{(2)}|\boldsymbol{\theta}_{-i}\right) &= \kappa_n^{(2)}\sum_x D(x,n)\beta_x^{(2)} - \sum_x E(x,n)\exp\left(\beta_x^{(1)} + \beta_x^{(2)}\kappa_n^{(2)} + \mathcal{R}(x,n)\right) \\
&\quad -\frac{1}{2\sigma_\kappa^2}\left(\kappa_n^{(2)} - \kappa_{n-1}^{(2)} - \delta\right)^2. \qquad (4.2.4)
\end{aligned}$$

In the case of the vector $\boldsymbol{\mathcal{R}}_1$ the density contains a single term from the conditional normal distribution, but also involves the term from the stationary distribution of $\boldsymbol{\mathcal{R}}_t$, which has been assigned to it in equation (4.1.5). Similarly for the last year n in the sample, $\boldsymbol{\mathcal{R}}_n$ includes only one term from the conditional normal distribution. For all the intermediate years $\boldsymbol{\mathcal{R}}_j$, $j \in (1,n)$, the pattern is identical with two conditionally normal terms involved. The full conditional log-posteriors of the vectors $\boldsymbol{\mathcal{R}}_t$ are given

as follows:

$$
\begin{aligned}
\pi\left(\boldsymbol{\mathcal{R}}_1|\boldsymbol{\theta}_{-i}\right) &= \sum_x \left\{ D(x,1)\mathcal{R}(x,1) - E(x,1)\exp\left(\beta_x^{(1)} + \beta_x^{(2)}\kappa_1^{(2)} + \mathcal{R}(x,1)\right) \right\} \\
&\quad -\frac{1}{2}\left[\left(\boldsymbol{\mathcal{R}}_2 - \mathcal{A}\boldsymbol{\mathcal{R}}_1\right)' V_Z^{-1}\left(\boldsymbol{\mathcal{R}}_2 - \mathcal{A}\boldsymbol{\mathcal{R}}_1\right) + \boldsymbol{\mathcal{R}}_1' V_{\mathcal{R}}^{-1}\boldsymbol{\mathcal{R}}_1 \right], \\
\pi\left(\boldsymbol{\mathcal{R}}_j|\boldsymbol{\theta}_{-i}\right) &= \sum_x \left\{ D(x,j)\mathcal{R}(x,j) - E(x,j)\exp\left(\beta_x^{(1)} + \beta_x^{(2)}\kappa_j^{(2)} + \mathcal{R}(x,j)\right) \right\} \\
&\quad -\frac{1}{2}\Bigg[\left(\boldsymbol{\mathcal{R}}_{j+1} - \mathcal{A}\boldsymbol{\mathcal{R}}_j\right)' V_Z^{-1}\left(\boldsymbol{\mathcal{R}}_{j+1} - \mathcal{A}\boldsymbol{\mathcal{R}}_j\right) \\
&\qquad + \left(\boldsymbol{\mathcal{R}}_j - \mathcal{A}\boldsymbol{\mathcal{R}}_{j-1}\right)' V_Z^{-1}\left(\boldsymbol{\mathcal{R}}_j - \mathcal{A}\boldsymbol{\mathcal{R}}_{j-1}\right) \Bigg], \text{with } 1 < j < n \\
\pi\left(\boldsymbol{\mathcal{R}}_n|\boldsymbol{\theta}_{-i}\right) &= \sum_x \left\{ D(x,n)\mathcal{R}(x,n) - E(x,n)\exp\left(\beta_x^{(1)} + \beta_x^{(2)}\kappa_n^{(2)} + \mathcal{R}(x,n)\right) \right\} \\
&\quad -\frac{1}{2}\left[\left(\boldsymbol{\mathcal{R}}_n - \mathcal{A}\boldsymbol{\mathcal{R}}_{n-1}\right)' V_Z^{-1}\left(\boldsymbol{\mathcal{R}}_n - \mathcal{A}\boldsymbol{\mathcal{R}}_{n-1}\right) \right]. \qquad (4.2.5)
\end{aligned}
$$

The vectors $\boldsymbol{\mathcal{R}}_t$ are sampled jointly for each fixed year t. There are three choices for proposing each vector for fixed time instance which generally affect the performance of the MCMC. All three choices involve a normal proposal with adaptive mean centred at the previous state of the chain. The difference between the potential proposal distribution options stems from the covariance matrix they possess. The simplest is to use a diagonal matrix and adjust the value of the involved parameter by trial and error. The second is to use an adaptive approximation of the posterior covariance matrix of $\boldsymbol{\mathcal{R}}_t$ using the observed Information Matrix, as was done for the period effects of the B-CBD model in Section 2.4.1. In this case, the combination of the exponential and quadratic form makes the calculation similarly simple in this case too. Finally, another choice is to use the latest estimate of matrix V_Z within the algorithm. Since this is the scheme's best estimate for matrix V_Z it also leads to reasonable results. The results presented here have been obtained by opting the second adaptive proposal distribution.

The residuals states are constrained in order to be made identifiable. The required relationships are:

$$
\sum_x \mathcal{R}(x,t) = \sum_x \beta_x^{(1)}\mathcal{R}(x,t) = \sum_x \beta_x^{(2)}\mathcal{R}(x,t) = 0, \forall t. \qquad (4.2.6)
$$

To apply the constraints, fix time t and denote $\widetilde{\mathcal{R}}(x,t)$ the adjusted residuals which satisfy relationships (4.2.6). Then:

$$\widetilde{\mathcal{R}}(x,t) = \mathcal{R}(x,t) + \phi_1 + \phi_2\beta_x^{(2)} + \phi_3\beta_x^{(1)}, \tag{4.2.7}$$

and $\boldsymbol{\phi} = (\phi_1, \phi_2, \phi_3)$ is given as:

$$\boldsymbol{\phi} = -\begin{pmatrix} m & \sum_x \beta_x^{(2)} & \sum_x \beta_x^{(1)} \\ \sum_x \beta_x^{(2)} & \sum_x (\beta_x^{(2)})^2 & \sum_x \beta_x^{(1)}\beta_x^{(2)} \\ \sum_x \beta_x^{(1)} & \sum_x \beta^{(1)}\beta_x^{(2)} & \sum_x (\beta_x^{(1)})^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_x \mathcal{R}(x,t) \\ \sum_x \beta_x^{(2)}\mathcal{R}(x,t) \\ \sum_x \beta_x^{(1)}\mathcal{R}(x,t) \end{pmatrix}.$$

By using equation (4.2.7) and applying appropriate adjustments for $\beta_x^{(2)}$ and $\kappa_t^{(2)}$ we ensure that equations (4.2.6) are satisfied.

The drift and the variance, δ and σ_κ^2, of the period effects model are conjugate, with full conditional posteriors given as:

$$\begin{aligned} \delta|\boldsymbol{\theta}_{-i} &\sim N\left(\frac{\sigma_\delta^2 \sum_{j=1}^{n-1}\left(\kappa_{j+1}^{(2)} - \kappa_j^{(2)}\right) + \delta_0\sigma_\kappa^2}{\sigma_\kappa^2 + (n-1)\sigma_\delta^2}, \frac{\sigma_\delta^2\sigma_\kappa^2}{\sigma_\kappa^2 + (n-1)\sigma_\delta^2} \right), \\ \sigma_\kappa^2|\boldsymbol{\theta}_{-i} &\sim IG\left(a + \frac{n-1}{2}, b + \frac{\sum_{j=1}^{n-1}\left(\kappa_{j+1}^{(2)} - \kappa_j^{(2)} - \delta\right)^2}{2} \right). \end{aligned} \tag{4.2.8}$$

Furthermore, the full conditional posterior of the autoregressive matrix $\mathcal{A}$ turns out to be the most complicated part of the model in terms of deriving the relevant distribution. The matrix is involved in the sum included at the ℓ_2 part of the likelihood but, moreover, it is involved in the system whose solution yields the matrix $V_\mathcal{R}$. Since no calculations are admitted for its specification, updates for them are based on normal random walk Metropolis steps centred at the current state of the chain and with appropriately tuned proposal variance. Lastly, the full conditional posterior distribution of the covariance matrix V_Z is as follows:

$$\begin{aligned} \pi\left(V_Z|\boldsymbol{\theta}_{-i}\right) &= -\frac{1}{2}\Big(\big(n_0 + m + n\big)\log\big(|V_Z|\big) + tr\big((M_\mathcal{R} + \Psi)V_Z^{-1}\big) \Big) \\ &\quad -\frac{1}{2}\Big(\log(|V_\mathcal{R}|) + \boldsymbol{\mathcal{R}}_1' V_\mathcal{R}^{-1}\boldsymbol{\mathcal{R}}_1 \Big), \end{aligned} \tag{4.2.9}$$

where $M_{\mathcal{R}}$ is the matrix constructed from the residuals of the residual process. That is,

$$M_{\mathcal{R}} = \sum_{j=1}^{n-1} \left[\Big(\boldsymbol{\mathcal{R}}_{j+1} - \mathcal{A}\boldsymbol{\mathcal{R}}_j \Big) \Big(\boldsymbol{\mathcal{R}}_{j+1} - \mathcal{A}\boldsymbol{\mathcal{R}}_j \Big)' \right]. \tag{4.2.10}$$

The first term of $\pi\left(V_Z|\boldsymbol{\theta}_{-i}\right)$ corresponds to an IW with $n+n_0-1$ degrees of freedom and scale matrix $\left(\mathcal{R}_{SM}+\Psi\right)^{-1}$. Although there is an additional term due to the assumption for the $\boldsymbol{\mathcal{R}}_1$ vector, in retrospect, the above distribution yields an efficient proposal distribution for the algorithm. This is mainly due to the low magnitude of the second part of the posterior relative to the magnitude of the term corresponding to the IW density. Hence, an independence MH step is utilised to update the covariance matrix V_Z.

The above discussion is now summarised in the following pseudo-algorithm:

Algorithm 4

Initialise the parameter vector

$$\boldsymbol{\theta}^{(0)} = \left\{ \left[\beta_x^{(1)}\right]^{(0)}, \left[\beta_x^{(2)}\right]^{(0)}, \left[\kappa_t^{(2)}\right]^{(0)}, \left[\boldsymbol{\mathcal{R}}_t\right]^{(0)}, \left[\mu\right]^{(0)}, \left[\sigma_\kappa^2\right]^{(0)}, \left[\mathcal{A}\right]^{(0)}, \left[V_Z\right]^{(0)} \right\}$$

and set the number of iterations, M.
For $j = 1, \ldots, M$

- For $i = 1, \ldots, m$
  - Sample and update $\left[\beta_i^{(1)}\right]^{(j)}$ from its full conditional distribution given by equation (4.2.2).
  - Sample a proposed value as $\widetilde{\beta}_i^{(2)} \sim N\left(\left[\beta_i^{(2)}\right]^{(j-1)}, \sigma^2_{\beta^{(2)}}\right)$
    - Calculate the corresponding acceptance ratio:
      $r = min\left\{1, \exp\left(\pi\Big(\widetilde{\beta}_i^{(2)}|\boldsymbol{\theta}_{-i}\Big) - \pi\Big(\left[\beta_i^{(2)}\right]^{(j-1)}|\boldsymbol{\theta}_{-i}\Big)\right)\right\}$
    - Simulate $u \sim U(0,1)$
    - If $r > u$, set: $\left[\beta_i^{(2)}\right]^{(j)} = \widetilde{\beta}_i^{(2)}$,
      otherwise set: $\left[\beta_i^{(2)}\right]^{(j)} = \left[\beta_i^{(2)}\right]^{(j-1)}$.
- For $i = 1, \ldots, n$
  - Sample a proposed value as $\widetilde{\kappa}_i^{(2)} \sim N\left(\left[\kappa_i^{(2)}\right]^{(j-1)}, \sigma^2_{\kappa^{(2)}}\right)$
    - Calculate the corresponding acceptance ratio:
      $r = min\left\{1, \exp\left(\pi\Big(\widetilde{\kappa}_i^{(2)}|\boldsymbol{\theta}_{-i}\Big) - \pi\Big(\left[\kappa_i^{(2)}\right]^{(j-1)}|\boldsymbol{\theta}_{-i}\Big)\right)\right\}$

* Simulate $u \sim U(0,1)$
* If $r > u$, set: $\left[\kappa_i^{(2)}\right]^{(j)} = \widetilde{\kappa}_i^{(2)}$,
  otherwise set: $\left[\kappa_i^{(2)}\right]^{(j)} = \left[\kappa_i^{(2)}\right]^{(j-1)}$.

- For $i = 1, \ldots, n$
  - Sample a proposed vector as $\widetilde{\boldsymbol{\mathcal{R}}}_i \sim N\left(\left[\boldsymbol{\mathcal{R}}_i\right]^{(j-1)}, \Sigma_{\boldsymbol{\mathcal{R}}_i}\right)$
    * Calculate the corresponding acceptance ratio:
      $r = min\left\{1, \exp\left(\pi\left(\widetilde{\boldsymbol{\mathcal{R}}}_i|\boldsymbol{\theta}_{-i}\right) - \pi\left(\left[\boldsymbol{\mathcal{R}}_i\right]^{(j-1)}|\boldsymbol{\theta}_{-i}\right)\right)\right\}$
    * Simulate $u \sim U(0,1)$
    * If $r > u$, set: $\left[\boldsymbol{\mathcal{R}}_i\right]^{(j)} = \widetilde{\boldsymbol{\mathcal{R}}}_i$,
      and apply the constraints (4.2.6) by implementing the equations:

$$\begin{aligned}
\left[\boldsymbol{\mathcal{R}}_i\right]^{(j)} &= [\boldsymbol{\mathcal{R}}_i]^{(j)} + \phi_1 + \phi_2 \left[\beta_x^{(2)}\right]^{(j)} + \phi_3 \left[\beta_x^{(1)}\right]^{(j)} \\
\left[\beta_x^{(1)}\right]^{(j)} &= \left[\beta_x^{(1)}\right]^{(j)} (1 - \phi_3) - \phi_1 \\
\left[\kappa_i^{(2)}\right]^{(j)} &= \left[\kappa_i^{(2)}\right]^{(j)} - \phi_2
\end{aligned}$$

    otherwise set: $\left[\boldsymbol{\mathcal{R}}_i\right]^{(j)} = [\boldsymbol{\mathcal{R}}_i]^{(j-1)}$.

- Apply the constraints (4.2.1) by implementing the equations:

$$\begin{aligned}
\left[\beta_x^{(1)}\right]^{(j)} &= \left[\beta_x^{(1)}\right]^{(j)} + \left[\beta_x^{(2)}\right]^{(j)} \times \left(\left[\kappa_1^{(2)}\right]^{(j)}\right) \\
\left[\kappa_t^{(2)}\right]^{(j)} &= \sum_x \left[\beta_x^{(2)}\right]^{(j)} \times \left(\left[\kappa_t^{(2)}\right]^{(j)} - \left[\kappa_1^{(2)}\right]^{(j)}\right) \\
\left[\beta_x^{(2)}\right]^{(j)} &= \left[\beta_x^{(2)}\right]^{(j)} \Big/ \left(\sum_x \left[\beta_x^{(2)}\right]^{(j)}\right).
\end{aligned}$$

- Sample for each element, or block of elements as described before, of the matrix $\mathcal{A}$ as $\widetilde{\mathcal{A}} \sim N\left(\left[\mathcal{A}\right]^{(j-1)}, \sigma_{\mathcal{A}}^2\right)$
  - Calculate the corresponding acceptance ratio:
    $r = min\left\{1, \exp\left(\pi\left(\widetilde{\mathcal{A}}|\boldsymbol{\theta}_{-i}\right) - \pi\left(\left[\mathcal{A}\right]^{(j-1)}|\boldsymbol{\theta}_{-i}\right)\right)\right\}$
  - Simulate $u \sim U(0,1)$
  - If $r > u$, set: $\left[\mathcal{A}\right]^{(j)} = \widetilde{\mathcal{A}}$,
    otherwise set: $\left[\mathcal{A}\right]^{(j)} = \left[\mathcal{A}\right]^{(j-1)}$.

- Sample for the parameters of the random walk, $[\delta]^{(j)}, [\sigma_\kappa^2]^{(j)}$, from their full conditional posteriors given in equations (4.2.8).

- Sample a proposed matrix from the proposal distribution with log-density $q_Z(\cdot)$ according to:

$$\widetilde{V_Z} \sim IW\Big((n+n_0-1),(\mathcal{R}_{SM}+\Psi)^{-1}\Big)$$

  - Calculate the corresponding acceptance ratio:
    $r = min\Bigg\{1, \exp\left(\pi\Big(\widetilde{V_Z}|\boldsymbol{\theta}_{-i}\Big)+q_Z\left([V_Z]^{(j-1)}\right)-\pi\Big([V_Z]^{(j-1)}\,|\boldsymbol{\theta}_{-i}\Big)-q_Z\left(\widetilde{V_Z})\right)\right)\Bigg\}$
  - Simulate $u \sim U(0,1)$
  - If $r > u$, set: $[V_Z]^{(j)} = \widetilde{V_Z}$,
    otherwise set: $[V_Z]^{(j)} = [V_Z]^{(j-1)}$.

The Bayesian paradigm has been employed to derive the full conditional log-posteriors for the latent states and the parameters of the BA-LC model. Those components have been utilised to illustrate the underlying MCMC algorithm used for simulating the posterior distribution of the full model.

4.3 Application

The BA-LC model is applied to the EW and Canada data and the results are analysed and compared to the MLE's of the original LC model. The algorithm for both populations runs for 2 million iterations, of which only every 80^{th} parameter vector is kept so that we end up with a sample of 25,000. The trace-plots shown in the graphs of this section are based on that sample, thinned at every 10^{th} of its entries for convenience. The cumulative trace-plots which are also plotted for some parameters of the BA-LC model are again based on the aforementioned sample, but after excluding an initial burn-in period, as indicated by the examination of the respective trace-plots. Attention is primarily paid to those components of the model that are essential for the forecasts. So, we are particularly interested in the convergence and summaries of the posterior distributions of the age effects, $\beta_x^{(1)}$ and $\beta_x^{(2)}$, the autoregression matrix $\mathcal{A}$, the covariance matrix V_Z, the δ and σ_κ^2 parameters of the random walk and the latent states, $\mathcal{R}(x,2009)$ and $\kappa_{2009}^{(2)}$, of the jump-off year of the data-set.

4.3.1 Convergence

The convergence of the chains is assessed graphically via their trace-plots and in some cases also via their cumulative mean and quantile plots. Figures 4.1, 4.2 and 4.3 show such plots for selected age effects, $\beta_x^{(1)}$ and $\beta_x^{(2)}$ for $x = 65, 75, 85$, the main autoregression coefficients of the matrix $\mathcal{A}$, the latest state, $\kappa_{2009}^{(2)}$, of the period effects and the parameters δ and σ_κ^2.

The chain converges in different rates for the various components of the parameter vector. For instance, examination of the traces for the age effects, $\beta_x^{(1)}$ and $\beta_x^{(1)}$, of the EW population in Figure 4.1 shows that if the random initial conditions are not close to valid posterior values, the chain is driven rapidly to the appropriate range. For example, this is the case for $\beta_{65}^{(1)}$ and $\beta_{85}^{(2)}$. On the other hand, if the initial value is close to the proper posterior range the individual components might wander around until the marginal chains settle, as is the case for $\beta_{85}^{(1)}$ and $\beta_{65}^{(2)}$. Moreover, in some cases the starting point of the chain may be close to representative posterior values, but the algorithm first explores possible values of the parameter space until it returns to the appropriate range. For example, this occurs with the age effect $\beta_{65}^{(2)}$ of the Canada data in Figure 4.2. Moreover, it is evident that the posterior distributions of $\beta_x^{(2)}$ indicate greater spread than those of $\beta_x^{(1)}$. In summary, the Gibbs components of the simulation scheme for the age effects $\beta_x^{(1)}$ and the MH steps for $\beta_x^{(2)}$ appear to provide sufficient approximations to the full posteriors of the respective parameters. The Gibbs algorithm also works particularly well for δ, but the scheme is more volatile for σ_κ^2. σ_κ^2 is assumed at its stationary point after the first 10,000 iterations for the EW population, and significantly earlier for Canada. The cumulative plots for the δ and σ_κ^2 in Figures 4.1 and 4.2 also denote the Poisson MLE's of the original LC model in red lines for the same data-sets. The convergent posterior sample of the drift has mean that is almost identical to the respective MLE's for both populations. On the the other hand, σ_κ^2 has notably lower posterior mean than the corresponding MLE's, systematically for both cases. This implies that the residuals process of the BA-LC model attracts some of the variability of the long-term dynamics of the original LC model.

The main elements of the matrix $\mathcal{A}$, α_4, α_5 and α_6, in Figure 4.3 show that in both populations the diffusion coefficients, α_4 and α_6, are strongly reverting close to zero and that they are attracted to those ranges rapidly. Parameter α_5, which is the cohort-type coefficient for the majority of ages in the data, appears to be the main driver for the evolution of the residual states. The cumulative plots seem satisfying enough for both data-sets. Interestingly, as will be seen in the next Section, the posteriors of α_4 and α_6 also have almost the same standard deviation in the case of the EW population, although this is not evident in their graphs.

Beyond aiding in the assessment of the algorithm the convergence analysis of this Section also assists in choosing the segment of the simulated sample that is summarised in the next subsection and is used for projecting the BA-LC model. Based on our findings, out of the sample of 25,000 simulated points we discard the first 10,000 as burn-in period.

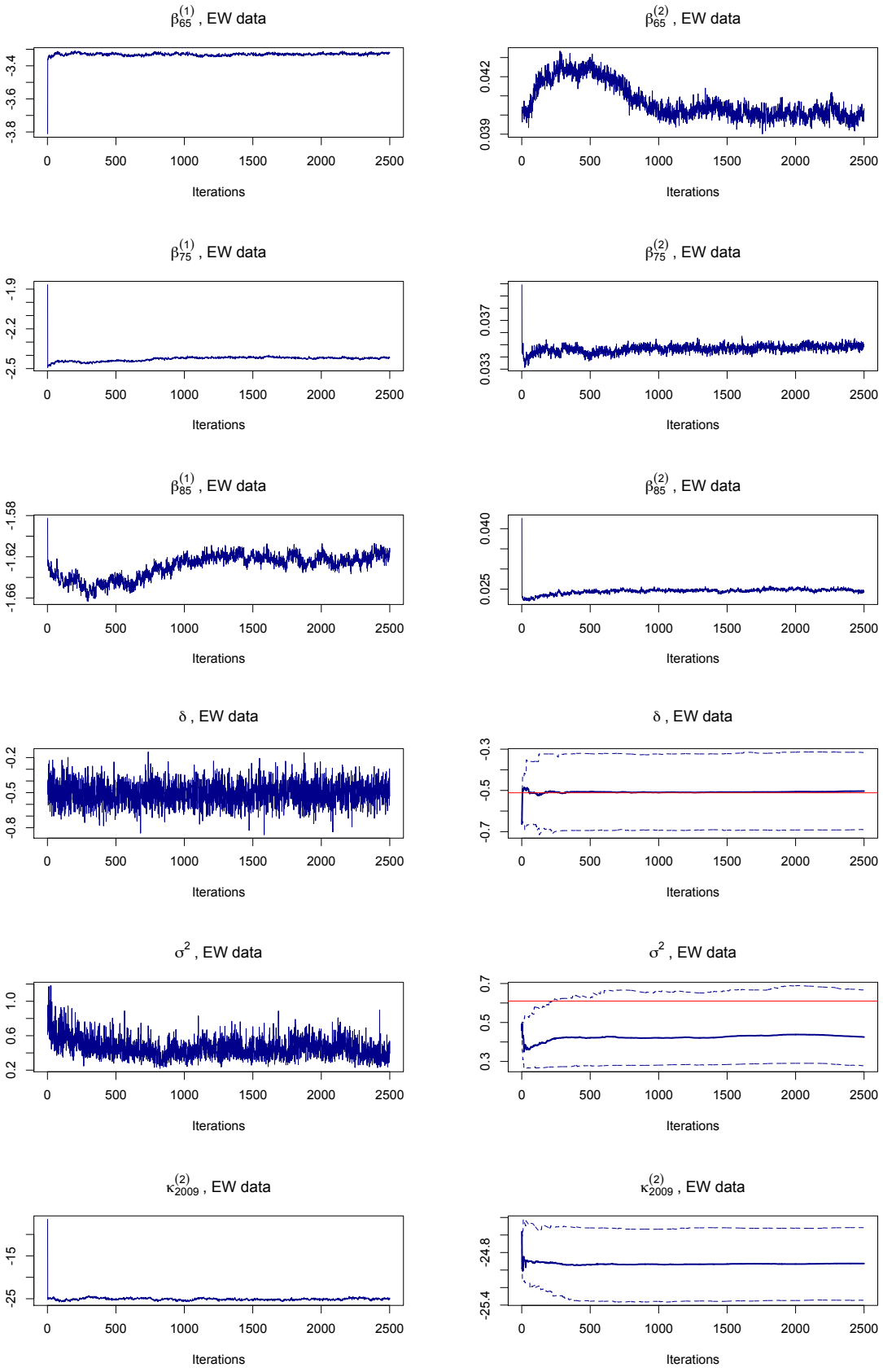

Figure 4.1: Trace-plots of age effects, $\beta_x^{(1)}$ and $\beta_x^{(2)}$, for $x = 65, 75$ and 85, of latent state $\kappa_{2009}^{(2)}$, and of drift and variance parameters, δ and σ_κ^2, of BA-LC model, EW data. The cumulative trace-plots of δ and σ_κ^2 also include the Poisson MLE's of the original LC model in red.

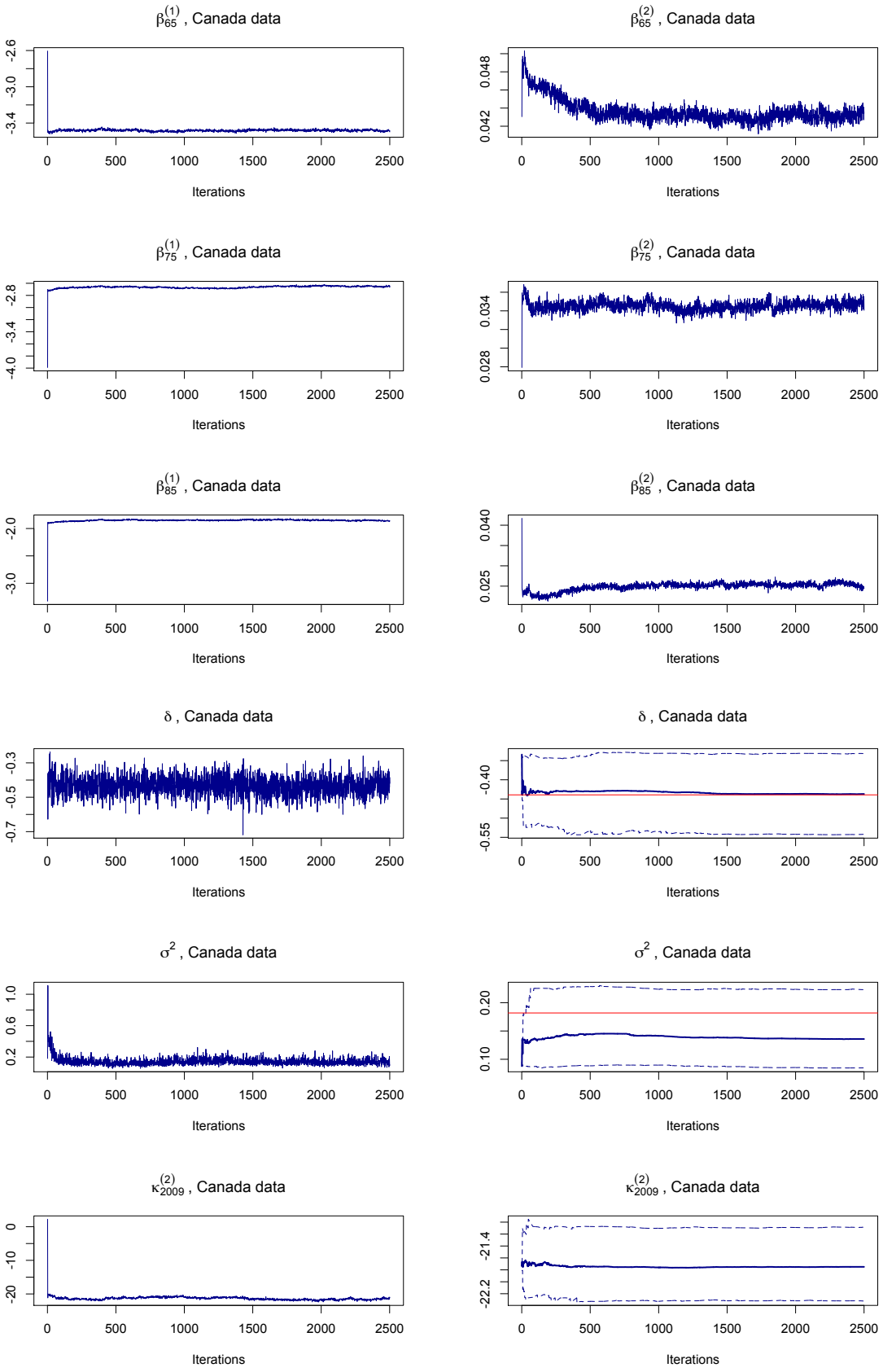

Figure 4.2: Trace-plots of age effects, $\beta_x^{(1)}$ and $\beta_x^{(2)}$, for $x = 65, 75$ and 85, of latent state $\kappa_{2009}^{(2)}$, and of drift and variance parameters, δ and σ_κ^2, of BA-LC model, Canada data. The cumulative trace-plots of δ and σ_κ^2 also include the Poisson MLE's of the original LC model in red.

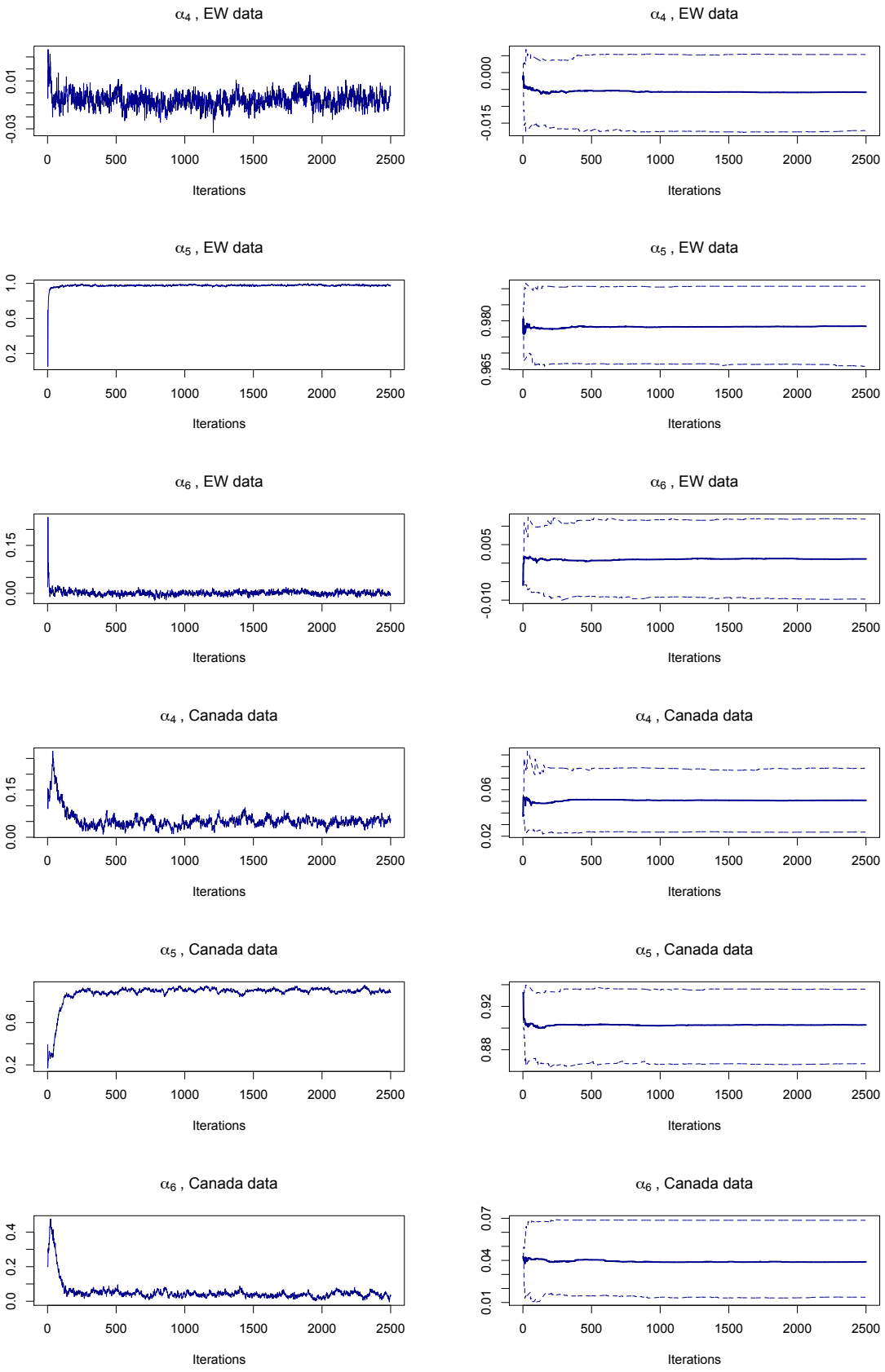

Figure 4.3: Trace-plots and cumulative trace-plots of autoregression parameters α_4, α_5 and α_6 of BA-LC model, EW and Canada data.

The remaining values are thinned at every 6^{th} iteration to yield a final sample of 2,500 estimates from the posterior distribution of the model. That is also the posterior sample that has been used to generate the cumulative plots of this subsection.

4.3.2 Posterior estimates

The summaries for the posterior distribution of the BA-LC model are presented here for the EW and Canada data-sets. The plots that accompany the analysis are also based on the convergent sample as discussed at the end of the previous subsection. Tables 4.1 and 4.2 show the means, standard deviations, 95% credibility intervals and p-values of Geweke's diagnostic for the marginal chains of the autoregression parameters of the matrix $\mathcal{A}$, the drift and variance of the random walk, δ and σ_κ^2, and the latest observed period effect $\kappa_{2009}^{(2)}$.

The autoregression coefficient for the first age in the data exhibits greater posterior mean for the Canada case, but has almost 2.5 times higher volatility in the EW data-set. The diffusion coefficient α_3 is negative in both cases, while it is significantly higher for the Canadian population compared to the EW data. As also noted in the previous Section, coefficients α_4 and α_6 converge very close to zero. For the EW data their means are equal in absolute value, but with negative signs, and have almost equal standard deviations. On the other hand, they are relatively close in mean for Canada, but the volatility of α_6 is almost 50% higher than that of α_4. Lastly, the main cohort related factor of the matrix $\mathcal{A}$, α_5, is greater for EW but quite comparable in magnitude to that of Canada.

	Mean	Std Deviation	L_{hpd}	U_{hpd}	p–value
α_1	0.89966224	0.05734858	0.76286843	0.98140771	0.442
α_2	1.14395811	0.12945437	0.92328486	1.43680483	0.130
α_3	−0.21071600	0.09667659	−0.39599291	−0.02383872	0.331
α_4	−0.00634030	0.00590943	−0.01804427	0.00512839	0.153
α_5	0.97787391	0.00613311	0.96596354	0.99022592	0.099
α_6	0.00066054	0.00573945	−0.01052340	0.01193191	0.558
$\kappa_{2009}^{(2)}$	−24.92139196	0.35397796	−25.50730257	−24.16475241	0.324
δ	−0.50198809	0.09244273	−0.68088767	−0.32798147	0.021
σ^2	0.43291101	0.09669059	0.27874423	0.65608987	0.759

Table 4.1: Summary of posterior distributions for the parameters of the BA-LC model and p-values of Geweke's convergence diagnostic, EW data.

The simulated sample from the posterior distribution of the period effect $\kappa_{2009}^{(2)}$ is of interest since it provides the potential starting points of the projection under the BA-LC model. If the MLE's of these latent states under the Poisson model are transformed to obey the constraints of the BA-LC model, their corresponding values are -25.06 and -21.53 for the EW and Canada data-sets, respectively. The means presented in Tables 4.1 and 4.2 are close to these previous estimates, while the credibility intervals cover them sufficiently well. Nevertheless, the posterior distribution of $\kappa_{2009}^{(2)}$ for EW has much greater spread than that of Canada. Although the variability of these starting points for projecting the period effects is not enough to reach estimates of $\kappa_t^{(2)}$ for previous years, the uncertainty introduced might have some impact upon the projections under the BA-LC model.

As it was also shown in the cumulative trace-plots of the previous subsection, the posterior means of δ are very close to the respective MLE's of the original LC model for both populations. The noteworthy point of the presented Tables is the almost twice greater level of uncertainty in the drift of the EW data compared to Canada. Conversely, the posterior distribution of σ_κ^2 is significantly downwards shifted compared to the Poisson MLE's for both data-sets. Thus, the BA-LC model appears to assign less variability in the innovations of the random walk model for the period dynamics, $\kappa_t^{(2)}$, as commented earlier.

	Mean	Std Deviation	L_{hpd}	U_{hpd}	p–value
α_1	0.94269944	0.01923330	0.90130786	0.97796786	0.193
α_2	0.95014308	0.09815505	0.79042233	1.14023132	0.518
α_3	-0.15614589	0.10604062	-0.37512283	0.04170243	0.177
α_4	0.04889206	0.01072536	0.02702702	0.06948247	0.429
α_5	0.90424366	0.01707023	0.87092807	0.93518558	0.239
α_6	0.03683169	0.01509330	0.01017996	0.06574454	0.598
$\kappa_{2009}^{(2)}$	-21.72858480	0.21941633	-22.14973560	-21.29314603	0.432
δ	-0.44179318	0.05284556	-0.54674114	-0.33750765	0.031
σ^2	0.13622161	0.03395450	0.08399088	0.21680804	0.020

Table 4.2: Summary of posterior distributions for the parameters of the BA-LC model and p-values of Geweke's convergence diagnostic, Canada data.

Finally, a comment is made on the performance of Geweke's diagnostic. The p-values of the marginal chains presented here, as well as those of the rest parameters of the model which are not explicitly tabulated, verify the satisfying convergence properties of the chosen posterior sample. Notably, parameter α_5 of the EW data produces

distinctively small p-value, but this is more due to the late settlement of the chain, as shown in Figure 4.3. Also, regarding the convergence properties of the marginal chains for the autoregressive parameters α_5 and α_6, noteworthy are the increasingly volatile cumulative percentiles in the case of the EW data, in contrast to the Canada population. Finally, the drift chains produce low p-values in both cases, which is surprising for the EW data-set considering the trace-plots of Figure 4.1.

The posterior distributions of the age effects, $\beta_x^{(1)}$ and $\beta_x^{(2)}$, and of the period factors, $\kappa_t^{(2)}$, are graphically summarised in Figure 4.4. The estimated means and 95% credibility percentiles are shown in solid and dotted lines, respectively. Common observations might be made for both data-sets. Firstly, the age effects $\beta_x^{(2)}$ develop the greatest uncertainty levels. On the contrary, the opposite is true for the age effects $\beta_x^{(1)}$ which have posterior distributions of very small spread. In each graph we also plot the original LC Poisson MLE's is blue solid lines, and the transformed sets of those parameters under the constraints of the BA-LC model. If the original MLE's of the LC model are denoted $\widetilde{\beta}_x^{(1)}$, $\widetilde{\beta}_x^{(2)}$ and $\widetilde{\kappa}_t^{(2)}$ we can move to the corresponding sets under the constraints of the BA-LC model, say $\beta_x^{(1)}$, $\beta_x^{(2)}$ and $\kappa_t^{(2)}$, through the relationships:

$$\begin{aligned} \kappa_t^{(2)} &= \widetilde{\kappa}_t^{(2)} - \widetilde{\kappa}_1^{(2)} \\ \beta_x^{(1)} &= \widetilde{\beta}_x^{(1)} + \widetilde{\beta}_x^{(2)}\widetilde{\kappa}_1^{(2)}, \end{aligned}$$

where $\widetilde{\kappa}_1^{(2)}$ is the entry for the first year of the data in the set of $\widetilde{\kappa}_t^{(2)}$.

The illustrated plots hint that the general trend of the estimates is maintained under both models. The explanation for this similarity is that the Poisson likelihood is the dominant component of the posterior distributions of the parameters, and since the $\mathcal{R}(x,t)$ states are not that prominent, the main shape of the estimates is maintained. The plots of $\beta_x^{(1)}$ show that the transformed original LC parameters fall very close to the posterior means of the BA-LC model. The observation is still valid for the period effects $\kappa_t^{(2)}$. However, in this case the transformed MLE's are evidently slightly greater for years in the middle of the data for both populations. Especially for the Canada data-set, the adjusted LC estimates fall significantly out of the 95% credibility percentiles for some of the years. Finally, in both cases the $\beta_x^{(2)}$ sets of parameters result in the same type of tilting pattern.

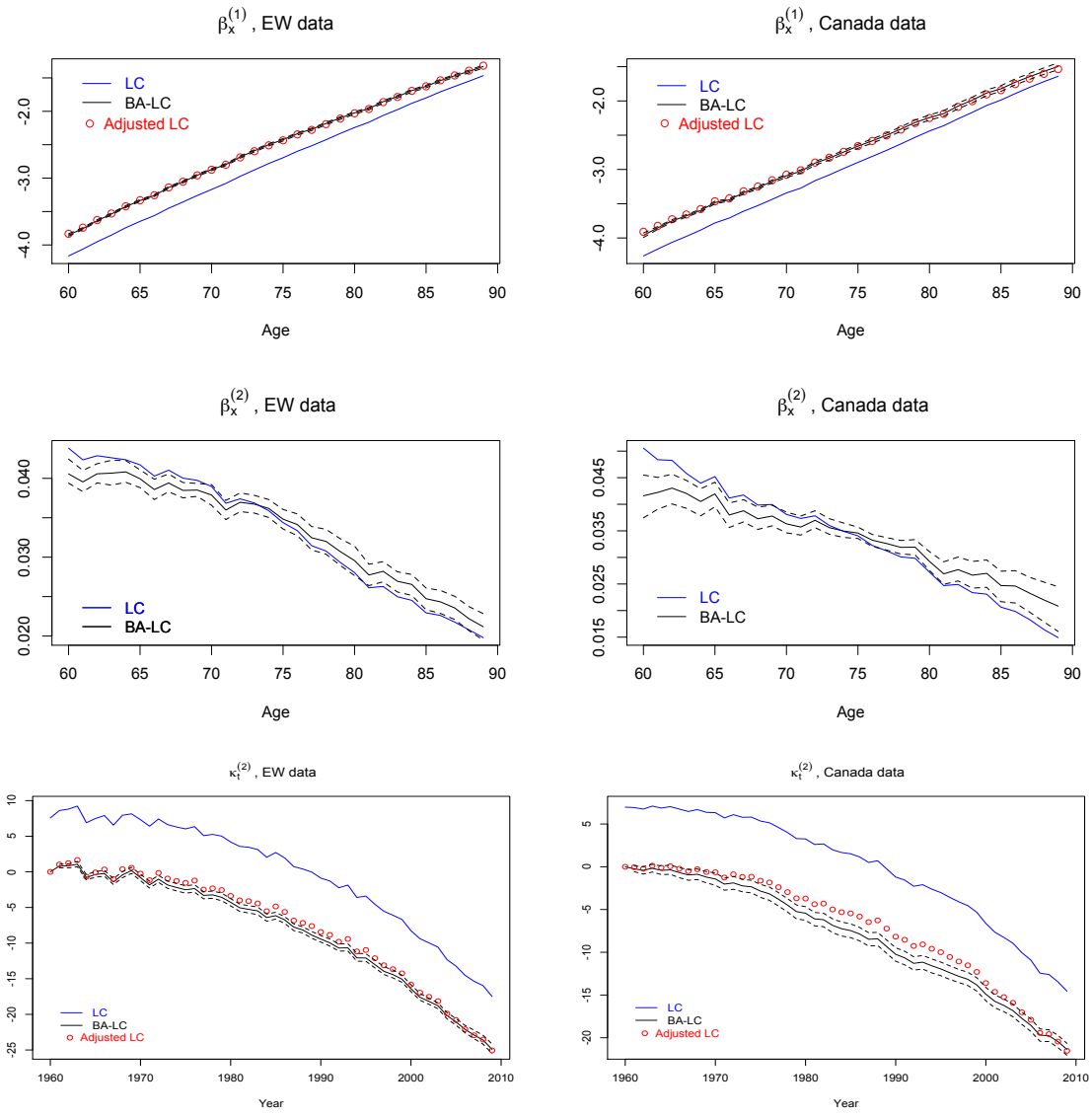

Figure 4.4: Posterior means and 95% credibility bands for $\beta_x^{(1)}$, $\beta_x^{(2)}$ and $\kappa_t^{(2)}$ of BA-LC model in black solid and dashed lines, respectively, for EW and Canada data. The original LC MLE's are plotted in blue and the adjusted LC estimates under the constraints of the BA-LC model are shown in red circles.

The latent residuals states of the BA-LC model serve a twofold purpose. Firstly, their samples assist in determining the distributions of the parameters of the underlying VAR model. Secondly, they incorporate missing structures of the LC modelling equation. The element-wise, across the convergent sample, posterior means of the $\mathcal{R}(x,t)$ states are shown in Figure 4.5. For both data-sets it is evident that the residuals process captures the cohort effect of the data, so that it compensates that weakness of the LC model. The contribution of $\mathcal{R}(x,t)$ in the log-death-rate is as high as 8% for some instances in the case of the EW data, but much lower when considering the Canada population.

Attention is also paid to the posterior distributions of the covariance matrices V_Z and $V_{\mathcal{R}}$. The former is a direct output of the algorithm, whereas the latter might be calculated given V_Z and the autoregression matrix $\mathcal{A}$ for any iteration. After calculating the element-wise mean of these model components the results are scaled to depict the correlation structures. Thus, in the middle panel of Figure 4.5 we have the measure of central tendency for the one-year ahead correlation of the innovations of the residuals process. In the bottom panel the corresponding statistic for the long-term correlation of the $\boldsymbol{\mathcal{R}}_t$ vectors is plotted for both datasets.

Matrix V_Z develops negative short-term correlation between the age band of 60-65 with ages of 70 and above for the EW population. The former age band is extended to 60-70 for the Canada population. Also, for both data-sets old ages are positively related, but this is much more intense for EW. Regarding matrix $V_{\mathcal{R}}$, both estimates imply some sort of diminishing long-term correlation structure amongst the entries of the $\boldsymbol{\mathcal{R}}_t$ vector. This form turns to negative between ages of the band 60-70 and above 80 for the EW data, although that negatively signed relationship is not observed in the Canada data. Lastly, it is noted that in the long-term, the residuals of ages at the bounds of the EW data-set seem strongly correlated to states of neighbouring ages. This is in contrast to the relationship developed for the residuals of middle ages of the EW data-set.

Subsequently, the distributions of the diagonal elements of matrix $V_{\mathcal{R}}$ are examined. Each these parameters corresponds to the stationary variance element of the vector $\boldsymbol{\mathcal{R}}_t$ across all ages of the data-set. The samples had been used for the calculation of the mean of matrix $V_{\mathcal{R}}$ previously, and are thus readily obtainable. These posteriors are summarised in the box-plots of Figure 4.6 and the results are substantially different between the two datasets. If the value 0.002 would be used as a threshold for numerical comparison between the two populations, the $V_{\mathcal{R}}$ diagonal element for age 60 of both of them is initially centred around that threshold. For the EW data, the residuals variance elements then develop distributions that are convexly shaped up to the age of 80, where they reach again the level of the threshold. Beyond age 80, there is an exponential increase between diagonal elements of the matrix $V_{\mathcal{R}}$ and age, with extrema of the residuals variance element for age 89 as great as 5 times the threshold. In contrast, the older ages' values for the Canada data are very close to the threshold. However, there is a similar, but rougher, convex form of relationship between the age and location of the posterior distribution of the corresponding diagonal element of $V_{\mathcal{R}}$.

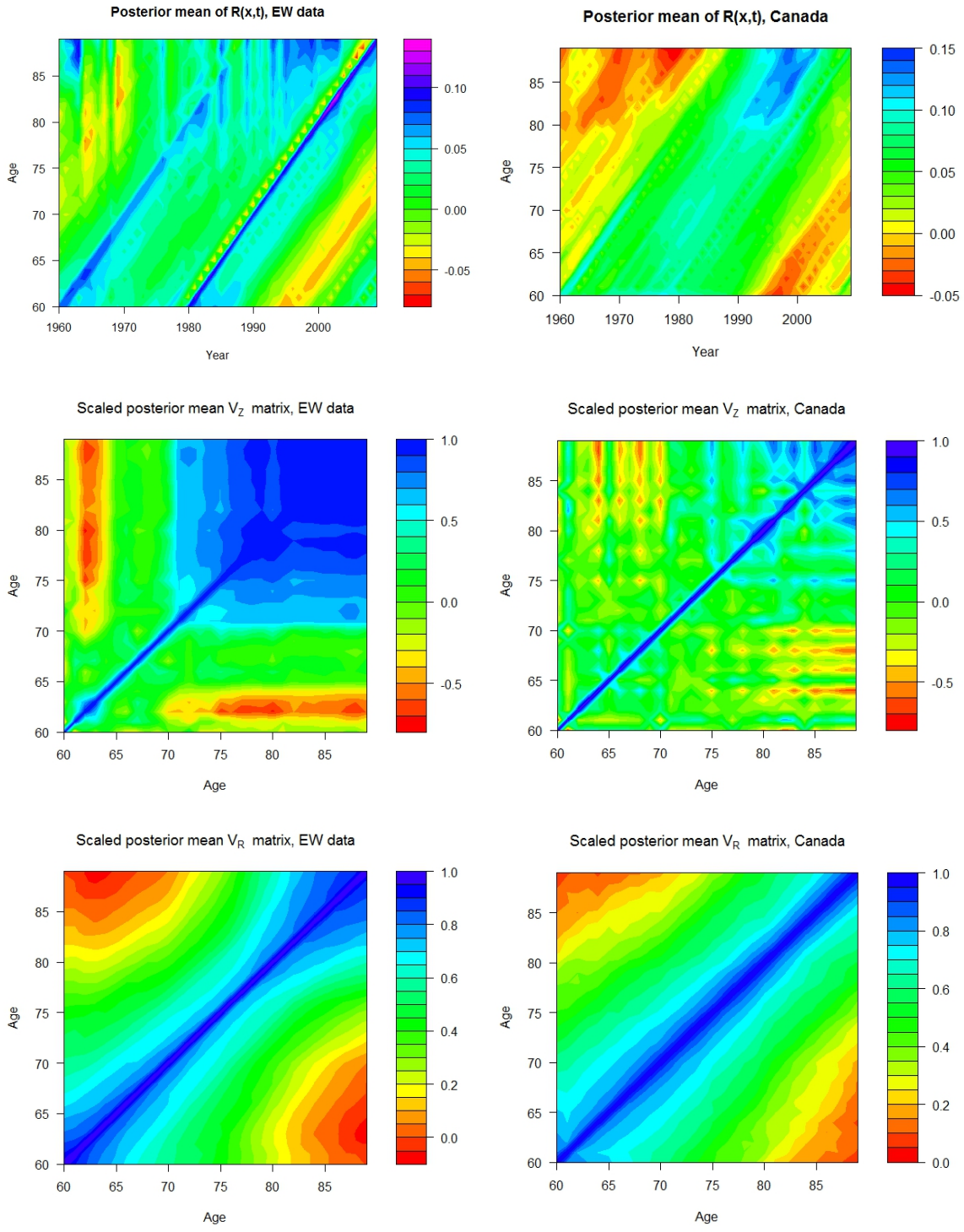

Figure 4.5: From top to bottom: Posterior means of residual sates, $\mathcal{R}(x,t)$
Posterior mean of matrix V_Z, scaled to represent correlations
Posterior mean of matrix $V_{\mathcal{R}}$, scaled to represent correlations, EW and Canada data left and right panels respectively.

The root of the previously demonstrated distinct behaviour for the diagonal elements of matrix $V_{\mathcal{R}}$ might be hidden in the posterior distributions of the diagonals of matrix V_Z. We thus now look at the distributions of these elements, whose means correspond to the unscaled values which were used to produced the matrices in the top panel of Figure 4.5. For the EW data, it appears that the distributions of the diagonal values of V_Z for ages above late 70's are centred around values that are up to 10 times greater relative to those of younger ages. There is a similar increasing relationship when considering the respective relationship of the Canada data, but it is not as great in magnitude as for EW. Additionally, the distributions of the diagonal elements of V_Z develop distributions with significant levels of decaying cross-correlation in the case of the EW data, with ages of the band 60-65 appearing to be negatively related to ages of 70 and above. In contrast, such structure is not observed in the case of the Canada population, which produces an unspecified structure of low values which fluctuate around zero between all ages of the dataset.

Conclusively, the increasing location of the distributions of the diagonal elements of matrix V_Z with age, along with the cross-correlation between the distributions of these entries for the EW data, result in the exponential increase that is observed in the location of the posterior distributions of the diagonal elements of matrix $V_{\mathcal{R}}$ in Figure 4.6.

The dependencies amongst the posterior distributions of the diagonal elements of matrix $V_{\mathcal{R}}$ are depicted via the graphs of their cross-correlation matrices in the bottom panel of Figure 4.6. The distributions show a strong positive relationship of values for ages from the start of the data-set up to age 75 for the EW population, which for ages above there diminishes. On the other hand, the cross-correlation matrix of the posterior distributions of the diagonal elements of $V_{\mathcal{R}}$ for the Canada data show a much more uniform structure.

Furthermore, we examine the cross-correlations between the parameters of the BA-LC model that are developed within the MCMC sampling. Firstly, we examine the bivariate scatter-plot between the parameters of the random walk model, δ and σ_κ^2, and conclude that there is nothing noticeable about their joint posterior distribution for both populations. Next, we examine the joint posterior distributions of the vectors of parameters $\beta_x^{(1)}$ and $\beta_x^{(2)}$ across all ages, and the bivariate distributions of the age effects for fixed age. Commonly for both data-sets, the joint posteriors of the age effect vectors across all ages appear positively correlated.

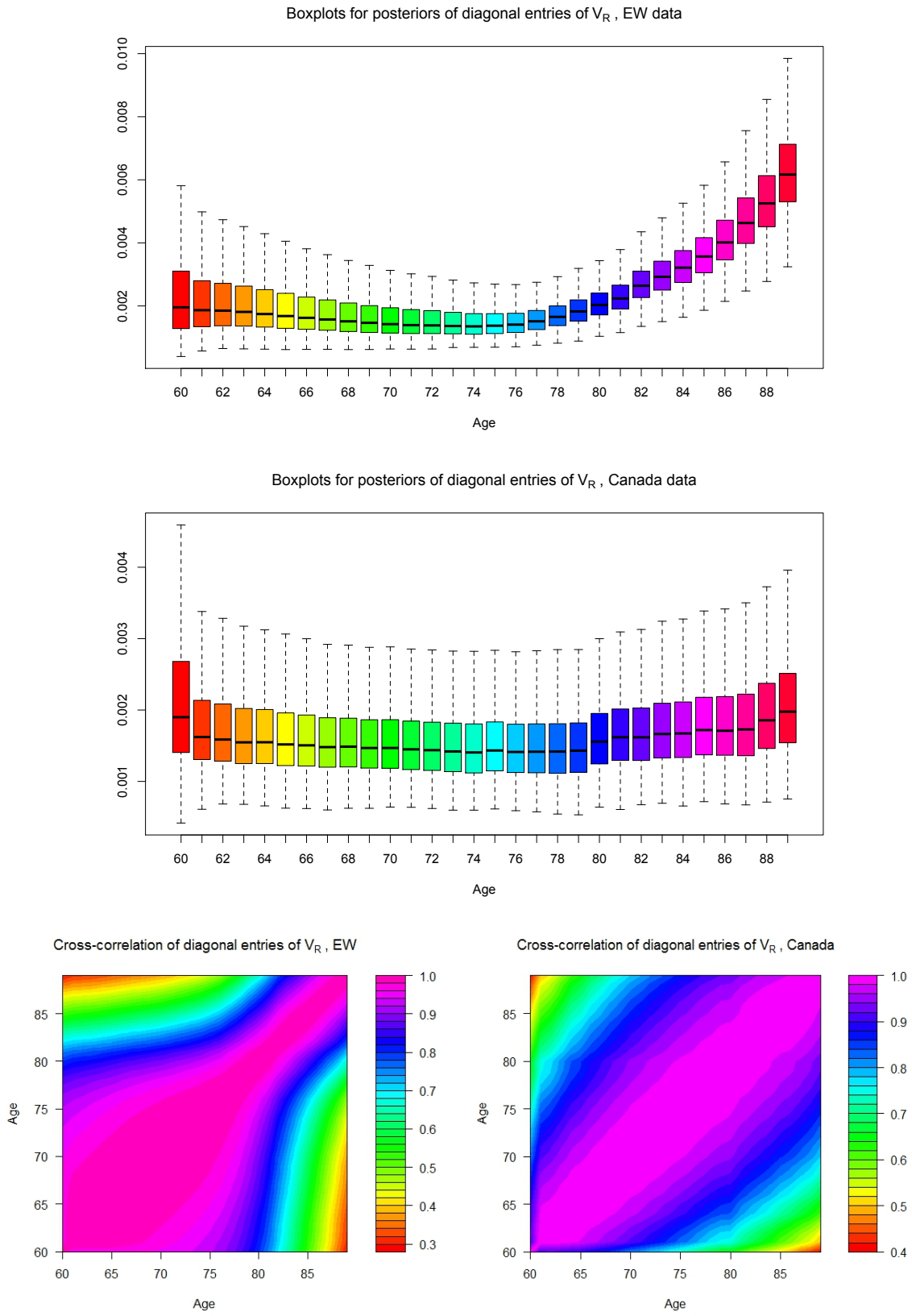

Figure 4.6: Box-plots for posteriors and posterior correlation matrices of diagonal entries of $V_{\mathcal{R}}$ matrix for EW and Canada data.

The result is more intense amongst younger and older ages of the data and mitigates when considering the middle age range. Also for both of the populations, the joint bivariate posteriors of the age effects for fixed age develop some limited evidence of positive dependence for ages at the bounds of the data-set, but no such relationship occurs for the majority of the ages under examination.

Finally, we examine the joint distributions of the autoregression parameters included within matrix $\mathcal{A}$. Figure 4.7 shows the scatter-plots of the pair of parameters (α_2, α_3) and of the triple $(\alpha_4, \alpha_5, \alpha_6)$. For both the EW and Canada populations, α_2 is strongly negatively correlated to α_3, which implies that the diffusion factor α_3 acts inversely to the cohort-type coefficient α_2. The relationship is also observed when considering the diffusion factors α_4 and α_6 against the cohort-type coefficient α_5 for the Canada data, but it is not evident for the EW data. Similar plots amongst all of the entries of matrix $\mathcal{A}$ show no other notable interdependencies that would have a natural interpretation.

We have analysed the results of the application of the BA-LC model to the EW and Canada data-sets. After assessing the convergence of the implemented algorithms, the resultant posterior distributions were summarised and their estimates were compared to Poisson MLE's of the original LC model. The additional components of the BA-LC model have offered useful insight in understanding the distinct dependence structures and sources of variability about the mortality trends of each population. For example, we inspect much greater levels of stationary variance for the latent $\boldsymbol{\mathcal{R}}_t$ states for higher ages in the EW data compared to Canada. Similarly, in the Canada data there exists a much stronger negative dependence between the diffusion factors α_4 and α_6 and the cohort-type coefficient α_5 of the autoregression matrix $\mathcal{A}$, compared to the EW data-set. On the other hand, there also exist systematic similarities between posterior distributions of the BA-LC for the two populations, such as the positive sort of posterior dependence amongst the entries of each of the period effects vectors.

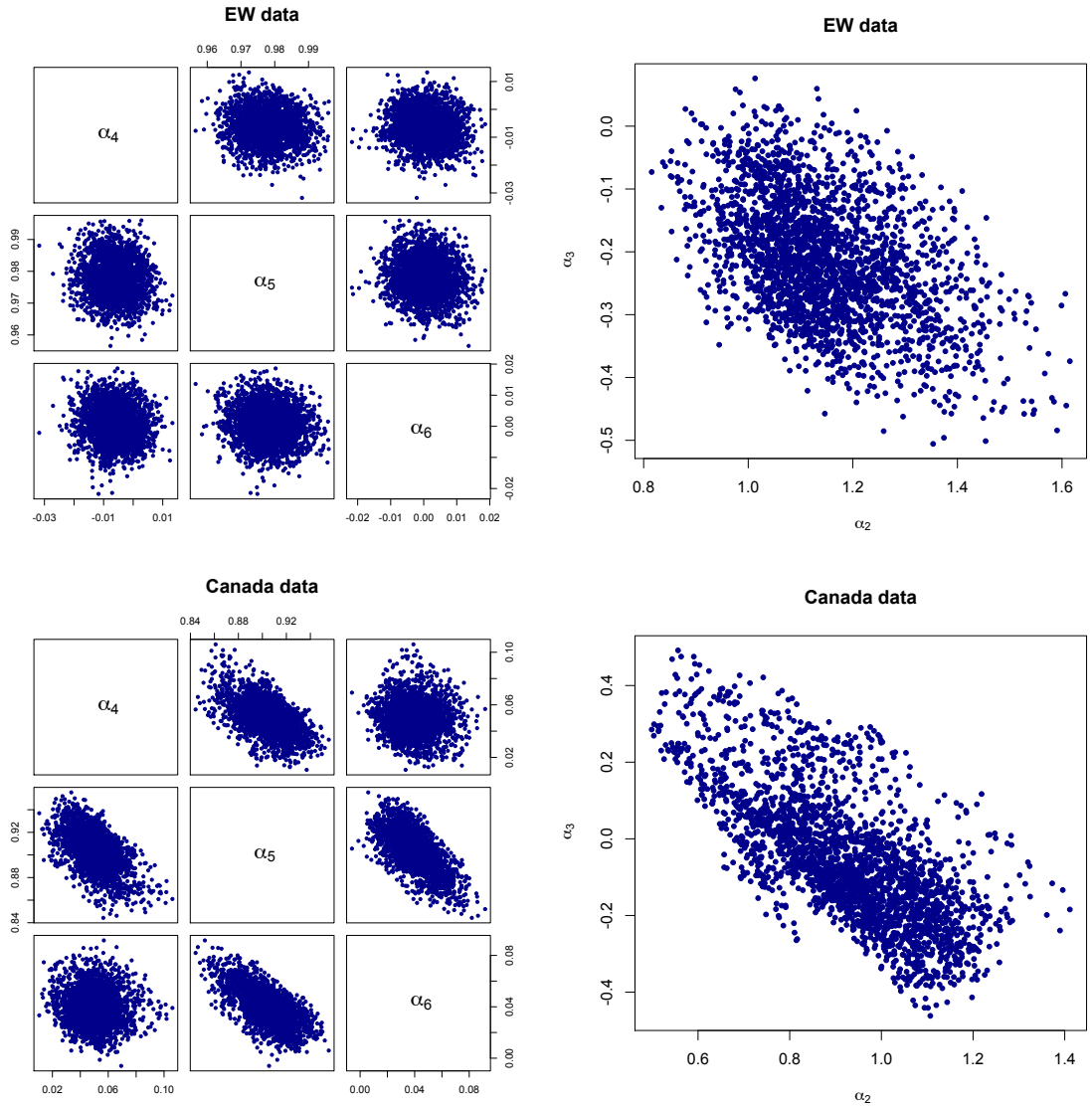

Figure 4.7: Bivariate scatterplots for posterior distributions of autoregression parameters, α_2 and α_3, and α_4, α_5 and α_6, of the VAR model for the residuals, $\mathcal{R}_t$, for EW and Canada data.

4.4 Summary

This Chapter presented the BA-LC model by combining the LC structure with the residuals model of Chapter 3 and by relaxing the assumption of conditionally independent residuals of the Poisson model for deaths. The resultant hierarchical model has more latent states than data points and hence estimation was favoured under the Bayesian framework of Chapter 2.

The long-term dynamics of the model are determined by the widespread random walk with drift choice for $\kappa_t^{(2)}$, while the latent residuals are connected so that they contribute local intercorrelations within the examined data-sets. The prior distribu-

tion of the full model includes an IW term which leads to a sufficient approximation for estimating the posterior distribution of the unstructured covariance matrix V_Z. However, the information supplied through that IW component is relatively vague. The remaining parameters specified in the prior are also not informative, apart from specifying the appropriate support for the elements of the autoregressive matrix, $\mathcal{A}$. The designed MCMC algorithm comprises a combination of MH steps, *e.g.* for the interaction and residual terms of the model, and Gibbs samplers, *e.g.* for the parameters of the random walk model, that yields a hybrid structure. After calibrating the model to the EW and Canada, convergence was illustrated for specific latent states and parameters of the model.

The posterior estimates of the model show that there are slight differences between the age and period effects of the original Poisson LC model and the BA-LC model, after the appropriate adjustments due to the different constraints employed. Parameters $\beta_x^{(2)}$ seem to be the most sensitive and those that exhibit the higher posterior uncertainty, whereas the variance of $\kappa_t^{(2)}$ is estimated to be lower compared to the Poisson MLE solution of the original LC model for both examined data-sets, though the drift estimate is matched between the two models for both populations. The residuals model is being accommodated smoothly within the suggested hierarchical structure. The latent residuals serve their purpose in compensating for the deviations of the LC fitted rates from the observations and the parameters of the VAR model maintain their interpretation. Most notably, the observation about the pattern of the diagonal elements of the covariance matrix V_Z in Section 3.3.2 is also depicted in the relevant posterior box-plots of Figure 4.6. Furthermore, the elements of the autoregression matrix, $\mathcal{A}$, show signs of posterior dependence, so that the diffusion coefficients α_2 for example act inversely to the cohort-type factors α_3. Finally, the impact of such structures varies for the two populations, with it being more intense in the case of Canada.

Chapter 5

Bayesian Augmented CBD Model

In this Chapter the latent residual states are accommodated under the CBD model to develop the BA-CBD model. The resultant hierarchical model is estimated by modifying slightly the algorithm of the B-CBD model in Chapter 2. The specification of the BA-CBD model is simplified by the single parameter diagonal structure of the covariance matrix V_Z. In this case, the long term dynamics of the model are driven by the vector of stochastic factors, $\boldsymbol{\kappa}_t$, of the original CBD model, and again all the properties of the original model are inherited to the augmented form too. Similarly to Chapter 4, we examine the convergence of the developed algorithms for selected latent states and parameters of the EW and Canada data, and compare the results to those of the CBD model under the conditionally independent Poisson model for deaths from Chapter 1.

5.1 Model description

The structure of the BA-CBD model is developed along similar lines to those of the BA-LC model. On the top level of the modelling hierarchy the deaths are still treated as Poisson variates, according to:

$$D(x,t)|m(x,t) \sim Poi\Big(E(x,t)m(x,t)\Big). \tag{5.1.1}$$

Although the CBD model focuses on the mortality rates, $q(x,t)$, and the binomial model for deaths developed in Chapter 1 might appear more intuitive, the use of the Poisson model, as in the case of the BA-LC model, allows us to compare the residuals augmentations of the two stochastic mortality structures within a unified framework under the same source of natural uncertainty.

The link equation gives the logit-mortality rate for age x last birthday, during calendar year t by:

$$\text{logit}\Big(q(x,t)\Big) = \kappa_t^{(1)} + \kappa_t^{(2)}\big(x - \bar{x}\big) + \mathcal{R}(x,t),$$

with $m(x,t)$ and $q(x,t)$ linked through equation (1.1.7). The random walk model suggested by Cairns et al. (2006a) is maintained for the period effect vectors, $\boldsymbol{\kappa}_t = \Big(\kappa_t^{(1)}, \kappa_t^{(2)}\Big)$, so that they are jointly modelled as:

$$\boldsymbol{\kappa}_{t+1} = \boldsymbol{\kappa}_t + \boldsymbol{\delta} + \boldsymbol{\zeta}_{t+1}, \tag{5.1.2}$$

where $\boldsymbol{\zeta}_t \sim N_2\big(0, V_\zeta\big)$ are *i.i.d.* bivariate normal random variables for all t, $\boldsymbol{\delta} = (\delta_1, \delta_2)$ is a constant drift vector and V_ζ is the covariance matrix of the disturbances $\boldsymbol{\zeta}_t$, parameterised as:

$$V_\zeta = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_2^2 \end{pmatrix}.$$

As was shown in Chapter 3, the bivariate underlying dynamics of the CBD model led to a distinctive form of correlation amongst the entries of the $\boldsymbol{\mathcal{R}}_t$ vector under the model of equation (3.3.4). Our empirical evidence, therefore, indicates that it may be possible to capture the pattern of the covariance matrix V_Z by a simple, for example diagonal, structure. Indeed, the empirical correlation matrix in Chapter 3 oscillates mostly between values within the interval (-0.1,0.1) and the implementation of the estimation algorithm indicates that such a diagonal specification is valid and is supported by both examined populations. Therefore, the set of distributional assumptions for the stochastic components of the BA-CBD model are given as follows:

$$\boldsymbol{\kappa}_{t+1}|\boldsymbol{\kappa}_t, \boldsymbol{\delta}, V_\zeta \sim N_2\big(\boldsymbol{\kappa}_t + \boldsymbol{\delta}, V_\zeta\big), \tag{5.1.3}$$

$$\boldsymbol{\mathcal{R}}_{t+1}|\boldsymbol{\mathcal{R}}_t, \mathcal{A}, V_Z \sim N_m\big(\mathcal{A}\boldsymbol{\mathcal{R}}_t, V_Z\big), \tag{5.1.4}$$

$$\boldsymbol{\mathcal{R}}_1|\mathcal{A}, V_Z \sim N_m\big(\mathbf{0}, V_{\mathcal{R}}\big). \tag{5.1.5}$$

Equation (5.1.3) refers to the conditionally normal density of the bivariate random walk dynamics from the model given in equation (5.1.2). Equations (5.1.4) and (5.1.5) correspond to the assumptions underlying the VAR residuals model, as for the BA-LC model. Finally, together with the Poisson assumption for deaths stated in equation (5.1.1), we have the full set of distributional assumptions about the BA-CBD model.

5.2 Model fitting

The fitting procedure of the presented model is now developed. The required components for the calibration of the BA-CBD are derived and the implemented stochastic simulation scheme is outlined.

The global parameter vector, $\boldsymbol{\theta}$, comprises the $\boldsymbol{\kappa}_t$ effects and the $\mathcal{R}_t$ terms, along with the parameters of the processes assigned to those components.
Therefore, let:

$$\boldsymbol{\theta} = \Big\{\boldsymbol{\kappa}_t, \boldsymbol{\mathcal{R}}_t, \boldsymbol{\delta}, V_\zeta, \mathcal{A}, V_Z\Big\}.$$

A similar MCMC algorithm is designed, as for the BA-LC model, to simulate from the full posterior distribution of the model. Since both models are implemented under the Poisson assumption for deaths, the retrospective comparison of the posterior distributions of the residuals models will be as valid and conclusive as possible.

5.2.1 Implementation

The likelihood of the BA-CBD model is based on the assumptions stated in the previous section. On the logarithmic scale, it is of the form:

$$\ell(\boldsymbol{\theta}) = \ell_1(\boldsymbol{\theta}) + \ell_2(\boldsymbol{\theta}) + \ell_3(\boldsymbol{\theta}) + C_1,$$

where C_1 is some constant independent of $\boldsymbol{\theta}$. For convenience and to emphasise the dependence on the parameter vector $\boldsymbol{\theta}$, let $u(\boldsymbol{\theta})$ denote the modelling equation:

$$u(\boldsymbol{\theta}) \equiv u(\boldsymbol{\theta}; x, t) = \text{logit}\Big(q(x,t)\Big) = \kappa_t^{(1)} + \kappa_t^{(2)}\big(x - \bar{x}\big) + \mathcal{R}(x,t).$$

The components ℓ_1 and ℓ_3 of the log-likelihood are given by the following equations:

$$\ell_1(\boldsymbol{\theta}) = \sum_{x,t}\left\{D(x,t)\log\left\{\log\Big[1+\exp\big(u(\boldsymbol{\theta})\big)\Big]\right\} - E(x,t)\log\Big[1+\exp\big(u(\boldsymbol{\theta})\big)\Big]\right\},$$

$$\ell_3(\boldsymbol{\theta}) = -\frac{1}{2}\left\{ \big(n-1\big) \log\big(|V_\zeta|\big) + \sum_{j=1}^{n-1} \Big[\big(\boldsymbol{\kappa}_{j+1} - \boldsymbol{\kappa}_j - \boldsymbol{\delta}\big)' V_\zeta^{-1} \big(\boldsymbol{\kappa}_{j+1} - \boldsymbol{\kappa}_j - \boldsymbol{\delta}\big) \Big] \right\}.$$

ℓ_1 comes from the Poisson assumption in terms of the parameterisation of the mortality rates, through equation (1.1.7). ℓ_3 corresponds to the conditional likelihood of the bivariate random walk model for the period effects, $\boldsymbol{\kappa}_t$. ℓ_2 is the same as in Section 4.2.1 since the residuals model is kept identical for both the BA-LC and BA-CBD models.

Finally, we set the prior specifications for the lowest level free parameters over all layers of hierarchy. Thus, a joint prior distribution is assigned for the components of the parameter sub-vector $\boldsymbol{\phi} = \Big\{\boldsymbol{\delta}, V_\zeta, \mathcal{A}, V_Z\Big\}$, which are assumed to be mutually independent.

The constant drift, $\boldsymbol{\delta}$, of the random walk model for the vector of period effects, $\boldsymbol{\kappa}_t$, is assigned a normal prior with mean $\boldsymbol{\delta}_0 = (0,0)'$ and covariance V_0, the 2×2 identity matrix. The prior is centred close to the maximum likelihood estimates of the original CBD model, but also diffuse enough and thus, not very restrictive. Matrix V_ζ is given the Jeffreys' uninformative prior (Cairns et al., 2006a), which is the most commonly employed uninformative option for the covariance matrix of the normal distribution (Gelman et al., 2003).

The prior for the autoregressive matrix $\mathcal{A}$ is again identical to that of Section 4.2.1 for the BA-LC model. Finally, the single parameter, v, of the diagonal matrix V_Z, is assigned an Inverse-Gamma (IG) prior distribution. The latter prior assignment is centred around a reasonable value, according to our empirical findings for the specific populations, but also with considerable spread adjusted through the coefficient of variation. This self-referential approach to setting the prior for v (Empirical Bayes) is discussed by Casella (1985). Given the low values for the diagonal entries of the empirical covariance matrix V_Z, such prior assignments assist in the convergence of the implemented algorithms.

The above discussion might be summarised in the following set of prior assumptions for the BA-CBD model:

$$\begin{aligned}
\boldsymbol{\delta} &\sim N(\boldsymbol{\delta}_0, V_0) \\
\alpha_1^* &\sim Beta(2,2) \quad \text{and} \quad \alpha_1 = 2\alpha_1^* - 1 \\
\alpha_3, \alpha_6 &\sim U(-1,1) \quad \text{and} \quad \alpha_2, \alpha_4, \alpha_5 \sim U(-\infty, +\infty) \\
v &\sim IG(a,b),
\end{aligned}$$

while the Jeffreys' prior for the covariance matrix V_ζ has density of the form:

$$p\left(V_\zeta\right) \propto |V_\zeta|^{-3/2},$$

and the hyper-prior parameters a and b of the IG distribution are determined as discussed above.

The corresponding joint log–prior density is given as:

$$\begin{aligned} p(\boldsymbol{\phi}) &= \log\left(1-\alpha_1^2\right) - \frac{1}{2}\left(\boldsymbol{\delta}-\boldsymbol{\delta}_0\right)' V_0^{-1}\left(\boldsymbol{\delta}-\boldsymbol{\delta}_0\right) \\ &\quad -(a+1)\log(v) - \frac{b}{v} - \frac{3}{2}\log\left(|V_\zeta|\right) + C_2, \end{aligned}$$

for $|\alpha_1|<1$, $|\alpha_3|<1$, $|\alpha_6|<1$, $v>0$ and V_0 positive definite. $p(\boldsymbol{\phi}) = -\infty$ otherwise, and C_2 is a normalising constant that does not depend on the parameters of interest.

The specification of the BA-CBD model in this Section allows the development of an MCMC algorithm to simulate from the full posterior distribution.

5.2.2 Posterior simulation

Exploiting the conjugate relationships of the model BA-CBD leads again to a hybrid MCMC algorithm in which the Gibbs sampler is employed whenever possible, and the MH scheme, or variants of it, is used when no analytic results occur.

The full conditional log-posteriors are now considered in turn for each component of the parameter vector $\boldsymbol{\theta}$. Similarly to Section 4.2.2 the conditioning in each case will be on the global parameter vector except for the parameter, or group of parameters, under consideration. The corresponding conditioning sub-vector is denoted $\boldsymbol{\theta}_{-i}$, where i is the parameter that is updated, while the other entries are fixed at their latest values.

The full conditional log-posteriors of the vector period effects, $\boldsymbol{\kappa}_t$ for fixed time t, are mixtures of Poisson and bivariate normal densities, given as:

$$\begin{aligned} \pi\left(\boldsymbol{\kappa}_1|\boldsymbol{\theta}_{-i}\right) &= \sum_x \left\{ D(x,1)\log\left\{\log\left[1+e^{\left(u(\boldsymbol{\kappa}_1)\right)}\right]\right\} - E(x,1)\log\left[1+e^{\left(u(\boldsymbol{\kappa}_1)\right)}\right]\right\} \\ &\quad -\frac{1}{2}\Big[\left(\boldsymbol{\kappa}_2-\boldsymbol{\kappa}_1-\boldsymbol{\delta}\right)' V_\zeta^{-1}\left(\boldsymbol{\kappa}_2-\boldsymbol{\kappa}_1-\boldsymbol{\delta}\right)\Big], \end{aligned}$$

$$\begin{aligned}
\pi\left(\boldsymbol{\kappa}_j|\boldsymbol{\theta}_{-i}\right) &= \sum_x \left\{ D(x,j)\log\left\{\log\left[1+e^{\left(u(\boldsymbol{\kappa}_j)\right)}\right]\right\} - E(x,j)\log\left[1+e^{\left(u(\boldsymbol{\kappa}_j)\right)}\right]\right\} \\
&\quad -\frac{1}{2}\Big[\left(\boldsymbol{\kappa}_{j+1}-\boldsymbol{\kappa}_j-\boldsymbol{\delta}\right)' V_\zeta^{-1}\left(\boldsymbol{\kappa}_{j+1}-\boldsymbol{\kappa}_j-\boldsymbol{\delta}\right) \\
&\qquad +\left(\boldsymbol{\kappa}_j-\boldsymbol{\kappa}_{j-1}-\boldsymbol{\delta}\right)' V_\zeta^{-1}\left(\boldsymbol{\kappa}_j-\boldsymbol{\kappa}_{j-1}-\boldsymbol{\delta}\right)\Big], \text{with } 1<j<n, \\
\pi\left(\boldsymbol{\kappa}_n|\boldsymbol{\theta}_{-i}\right) &= \sum_x \left\{ D(x,n)\log\left\{\log\left[1+e^{\left(u(\boldsymbol{\kappa}_n)\right)}\right]\right\} - E(x,n)\log\left[1+e^{\left(u(\boldsymbol{\kappa}_n)\right)}\right]\right\} \\
&\quad -\frac{1}{2}\Big[\left(\boldsymbol{\kappa}_n-\boldsymbol{\kappa}_{n-1}-\boldsymbol{\delta}\right)' V_\zeta^{-1}\left(\boldsymbol{\kappa}_n-\boldsymbol{\kappa}_{n-1}-\boldsymbol{\delta}\right)\Big].
\end{aligned}$$

The functions of the vectors $\boldsymbol{\kappa}_1$ and $\boldsymbol{\kappa}_n$ include a single term coming from the Normal density. For period effects vectors for intermediate years, $\boldsymbol{\kappa}_j$, there are two such terms involved. The latent states $\boldsymbol{\kappa}_t$ are sampled jointly for fixed year t via the approximation used in Section 2.4.1 for the Poisson B-CBD model. The tuning constant for both data-sets is set to $c = 2.8$.

Similarly, the time series structure of the model for $\boldsymbol{\mathcal{R}}_t$ yields distinct full conditional posterior densities for the years at the bounds of the dataset compared to intermediate years. The comments of Section 4.2.2 about the full conditional log-posterior of the vectors $\boldsymbol{\mathcal{R}}_t$ apply identically here as well. The diagonal matrix V_Z has only qualitative difference and causes no practical distinction in terms of the solution of the required system given in equation (3.3.9) of Section 3.3. The respective densities follow closely those derived in the BA-LC model, but with different form for their likelihood component. For all distinct cases, they are given as follows:

$$\begin{aligned}
\pi\left(\boldsymbol{\mathcal{R}}_1|\boldsymbol{\theta}_{-i}\right) &= \sum_x \left\{ D(x,1)\log\left\{\log\left[1+e^{\left(u(\boldsymbol{\mathcal{R}}_1)\right)}\right]\right\} - E(x,1)\log\left[1+e^{\left(u(\boldsymbol{\mathcal{R}}_1)\right)}\right]\right\} \\
&\quad -\frac{1}{2}\Big[\left(\boldsymbol{\mathcal{R}}_2-\mathcal{A}\boldsymbol{\mathcal{R}}_1\right)' V_Z^{-1}\left(\boldsymbol{\mathcal{R}}_2-\mathcal{A}\boldsymbol{\mathcal{R}}_1\right)+\boldsymbol{\mathcal{R}}_1' V_{\mathcal{R}}^{-1}\boldsymbol{\mathcal{R}}_1\Big], \\
\pi\left(\boldsymbol{\mathcal{R}}_j|\boldsymbol{\theta}_{-i}\right) &= \sum_x \left\{ D(x,j)\log\left\{\log\left[1+e^{\left(u(\boldsymbol{\mathcal{R}}_j)\right)}\right]\right\} - E(x,j)\log\left[1+e^{\left(u(\boldsymbol{\mathcal{R}}_j)\right)}\right]\right\} \\
&\quad -\frac{1}{2}\Big[\left(\boldsymbol{\mathcal{R}}_{j+1}-\mathcal{A}\boldsymbol{\mathcal{R}}_j\right)' V_Z^{-1}\left(\boldsymbol{\mathcal{R}}_{j+1}-\mathcal{A}\boldsymbol{\mathcal{R}}_j\right) \\
&\qquad +\left(\boldsymbol{\mathcal{R}}_j-\mathcal{A}\boldsymbol{\mathcal{R}}_{j-1}\right)' V_Z^{-1}\left(\boldsymbol{\mathcal{R}}_j-\mathcal{A}\boldsymbol{\mathcal{R}}_{j-1}\right)\Big], \text{with } 1<j<n,
\end{aligned}$$

$$\pi\left(\boldsymbol{\mathcal{R}}_n|\boldsymbol{\theta}_{-i}\right) = \sum_x \left\{ D(x,n)\log\left\{\log\left[1+e^{\left(u(\boldsymbol{\mathcal{R}}_n)\right)}\right]\right\} - E(x,n)\log\left[1+e^{\left(u(\boldsymbol{\mathcal{R}}_n)\right)}\right]\right\}$$

$$-\frac{1}{2}\Big[\left(\boldsymbol{\mathcal{R}}_n - \mathcal{A}\boldsymbol{\mathcal{R}}_{n-1}\right)' V_Z^{-1}\left(\boldsymbol{\mathcal{R}}_n - \mathcal{A}\boldsymbol{\mathcal{R}}_{n-1}\right)\Big].$$

The residuals of the BA-CBD model are constrained similarly to the BA-LC model. In this case the constraints are:

$$\sum_x \mathcal{R}(x,t) = \sum_x (x-\bar{x})\mathcal{R}(x,t) = 0, \forall t. \tag{5.2.1}$$

As with the BA-LC model, equation

$$\widetilde{\mathcal{R}}(x,t) = \mathcal{R}(x,t) + \phi_1 + \phi_2(x-\bar{x}), \tag{5.2.2}$$

along with corresponding adjustments for $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ ensure that constraints (5.2.1) are satisfied. Coefficients ϕ_1 and ϕ_2 are given as:

$$\phi_1 = -\frac{1}{m}\sum_x \mathcal{R}(x,t) \text{ and } \phi_1 = -\frac{1}{\sum_x (x-\bar{x})^2}\sum_x \mathcal{R}(x,t)(x-\bar{x}). \tag{5.2.3}$$

The parameters of the random walk, $\boldsymbol{\delta}$ and V_ζ, are conjugate with tractable full conditional posteriors again as in the B-CBD model of Chapter 2:

$$\begin{aligned} V_\zeta|\boldsymbol{\theta}_{-i} &\sim IW\left(n-2, \frac{1}{n-1}\widehat{V}\right), \\ \boldsymbol{\delta}|\boldsymbol{\theta}_{-i} &\sim N\left(\boldsymbol{\delta}_p, V_p\right), \end{aligned} \tag{5.2.4}$$

and the required parameters are taken from the set of equations 2.4.3. The simulation of the autoregressive matrix $\mathcal{A}$ again follows exactly as in Section 4.2.2. Finally, the full conditional log-posterior density of the parameter v on the diagonal of the covariance matrix V_Z is given by:

$$\pi\left(v|\boldsymbol{\theta}_{-i}\right) = -\log\left(v\right)\left(\frac{m\left(n-1\right)}{2} + a + 1\right) - \frac{1}{v}\left(b + \frac{tr\left(M_{\mathcal{R}}\right)}{2}\right)$$

$$-\frac{1}{2}\left(\log\left(|V_{\mathcal{R}}|\right) + \boldsymbol{\mathcal{R}}_1' V_{\mathcal{R}}^{-1}\boldsymbol{\mathcal{R}}_1\right),$$

where $tr(\cdot)$ indicates the trace of a matrix, and $M_{\mathcal{R}}$ is as in equation (4.2.10). The density $\pi\left(v|\boldsymbol{\theta}_{-i}\right)$ is that of an exact IG distribution. However, its specification is impossible due to the complexity of the stationary covariance matrix $V_{\mathcal{R}}$. Nevertheless, the sum of the first two terms of $\pi\left(v|\boldsymbol{\theta}_{-i}\right)$ provides a sufficient and tractable IG

approximation of the true target distribution. Similarly to the sampling of the IW covariance matrices of the BA-LC model, the approximate IG distribution is used as the proposal distribution for the MH step implemented to update that parameter.

The algorithm for the estimation of the BA-CBD model is now developed. Since sampling for the residuals model is identical as for the BA-LC algorithm, some steps are just referenced.

Algorithm 5

Initialise the parameter vector

$$\boldsymbol{\theta}^{(0)} = \left\{ \boldsymbol{\kappa}_t^{(0)}, \boldsymbol{\mathcal{R}}_t^{(0)}, \boldsymbol{\delta}^{(0)}, V_\zeta^{(0)}, \mathcal{A}^{(0)}, v^{(0)} \right\},$$

and set the number of iterations, M.
For $j = 1, \ldots, M$

- For $i = 1, \ldots, n$
  - Sample a candidate period effects vector form the proposal distribution, with density $q_{\boldsymbol{\kappa}_i}$, as:

$$\boldsymbol{\kappa}_i^* \sim N\left(\boldsymbol{\kappa}_i^{(j-1)}, c\mathcal{I}\left(\boldsymbol{\kappa}_i^{(j-1)} \right) \right) \tag{5.2.5}$$

  - Calculate the acceptance ratio according to:

$$r\left(\boldsymbol{\kappa}_i^{(j-1)}, \boldsymbol{\kappa}_i^* \right) = min \left\{ 1, \frac{q_{\boldsymbol{\kappa}_i}\left(\boldsymbol{\kappa}_i^{(j-1)} | \boldsymbol{\kappa}_i^* \right) \times \exp\left(\pi\left(\boldsymbol{\kappa}_i^* | \boldsymbol{\theta}_{-i} \right) \right)}{q_{\boldsymbol{\kappa}_i}\left(\boldsymbol{\kappa}_i^* | \boldsymbol{\kappa}_i^{(j-1)} \right) \times \exp\left(\pi\left(\boldsymbol{\kappa}_i^{(j-1)} | \boldsymbol{\theta}_{-i} \right) \right)} \right\}.$$

  - Generate $U \sim U(0,1)$.
  - If $U \leq r\left(\boldsymbol{\kappa}_i^{(j-1)}, \boldsymbol{\kappa}_i^* \right)$, set $\boldsymbol{\kappa}_i^{(j)} = \boldsymbol{\kappa}_i^*$, otherwise set $\boldsymbol{\kappa}_i^{(j)} = \boldsymbol{\kappa}_i^{(j-1)}$.
  - Sample the $\mathcal{R}_t$ states as described in **Algorithm 4**, and if accepted apply the constraints of equation (5.2.1). .
- Sample the autoregressive matrix $\mathcal{A}$ as described in **Algorithm 4**.
- Sample the drift $[\boldsymbol{\delta}]^{(j)}$, and the covariance matrix $[V_\zeta]^{(j)}$, from their full conditional posterior distributions given in the set of equations (5.2.4).
- Sample a candidate variance parameter from the proposal distribution, with density q_v, as:

$$v^* \sim IG\left(\left(\frac{m(n-1)}{2} + a + 1 \right), \left(b + \frac{tr(\mathcal{R}_{SM})}{2} \right) \right).$$

– Calculate the acceptance ratio according to:

$$r\big(v^{(j-1)}, v^*\big) = min\Bigg\{1, \frac{q_v\Big(v^{(j-1)}\Big) \times \exp\Big(\pi\big(v^*|\boldsymbol{\theta}_{-i}\big)\Big)}{q_v\big(v^*\big) \times \exp\Big(\pi\big(v^{(j-1)}|\boldsymbol{\theta}_{-i}\big)\Big)}\Bigg\}.$$

– Generate $U \sim U(0,1)$.

– If $U \leq r\big(v^{(j-1)}, v^*\big)$, set $v^{(j)} = v^*$, otherwise set $v^{(j)} = v^{(j-1)}$.

- Save the iteration and continue if $j < M$.

Iterate the described scheme until convergence yields a sample from the marginal posterior distributions of all the unknowns.

5.3 Application

The BA-CBD model is applied to the EW and Canada data and the results are analysed and compared to the MLE's of the original CBD model and of model M7. The identity of the algorithms is the same as that of those in Section 4.3, so that the posterior sample is of size 25,000. The cumulative trace-plots are similarly based on the full sample after excluding the burn-in period. For the BA-CBD model we focus on the drift vector, $\boldsymbol{\delta}$, the covariance matrix, V_ζ, the autoregression matrix $\mathcal{A}$, the single parameter v of the matrix V_Z and the latent states, $\mathcal{R}(x, 2009)$ and $\boldsymbol{\kappa}_{2009}$, of the final year in the data-set.

5.3.1 Convergence

The convergence properties of the implemented algorithms are assessed via the trace-plots and cumulative trace-plots of the marginal chains, as in Section 4.3.1 for the BA-LC model. Figures 5.1 and 5.2 illustrate the cumulative trace-plots of the main autoregression coefficients, α_4, α_5 and α_6, of the drift vector $\boldsymbol{\delta}$ and of the variance parameter v of matrix V_Z for the EW and Canada populations, respectively.

After examining the corresponding trace-plots, the simulation scheme seems to be converging faster for the BA-CBD model compared to the BA-LC. Hence, the cumulative trace-plots of Figures 5.1 and 5.2 are based on the sample of the 25,000 simulated values after discarding the first 5,000 iterations as burn-in period and thinning out the resultant chains at every 8^{th} iteration. The results of Geweke's diagnostic, shown in the tables of the following section, are applied to that convergent segment of the posterior sample and broadly support the convergence of the involved parameters.

The Gibbs parts of the MCMC for the parameters of the random walk model perform adequately, so that the marginal chains of the components of the vector $\boldsymbol{\delta}$ and of the matrix V_ζ are located close to their posterior distributions almost since the first few iterations. The red lines in the cumulative trace-plots of the drifts in graphs indicate the Poisson MLE's under the original CBD model. For both populations, the posterior mean of δ_1 is very close to the respective MLE. This implies that the impact of the auxiliary residuals in the estimation of the level of the mortality surface, that is parameters $\kappa_t^{(1)}$, has limited effect in the trend they develop, since the posterior distribution of $\boldsymbol{\delta}$ depends on the trend of the latent period states. The convergent mean of δ_2 seems to be slightly shifted downwards for the EW data. However, in the case of the Canada population, the cumulative plots indicate that convergence is achieved later in that case. Parameter σ_1^2 is very close to the Poisson MLE's for both data-sets. The same is true for σ_2^2 in the case of the EW data, but σ_2^2 converges to a much lower value than the corresponding Poisson MLE for Canada. The identical observation is true for parameter σ_{12} regarding the convergence level between the two populations.

The IG approximation for parameter v yields acceptance ratio over 50% within the stochastic simulation. The effective sampling for v is evident in both cases. The distant starting value of the EW data implementation is corrected within a few cycles until convergence to the range of the posterior distribution. The Canada scheme initiates from a state close to the posterior distribution and never leaves that range for the following 2 million iterations. The very small range of values for v imply negligible variability of the residuals state vector $\boldsymbol{\mathcal{R}}_t$. The convergence seems to be stable for both populations and the cumulative trace-plots of the convergent segment of the MCMC sample emphasise the quality of the posterior distribution of v.

The MH components of the algorithm function sufficiently too. If the initial conditions are far away from the proper posterior range, as for example for α_5 and α_6 of both data-sets, the chain reaches the correct space within a few thousand real iterations and spends the remaining simulation time therein. The cumulative trace-plots of the convergent part of the MCMC sample are more stable for the EW case. Notably, coefficient α_5 of the Canadian population seems to have significantly lower mean than that of EW, although that difference in the cohort-type coefficient was not observed for the BA-LC model between the two populations. Similar comments regarding the convergence properties of the marginal chains apply for the latent states, $\boldsymbol{\kappa}_t$ and $\boldsymbol{\mathcal{R}}_t$.

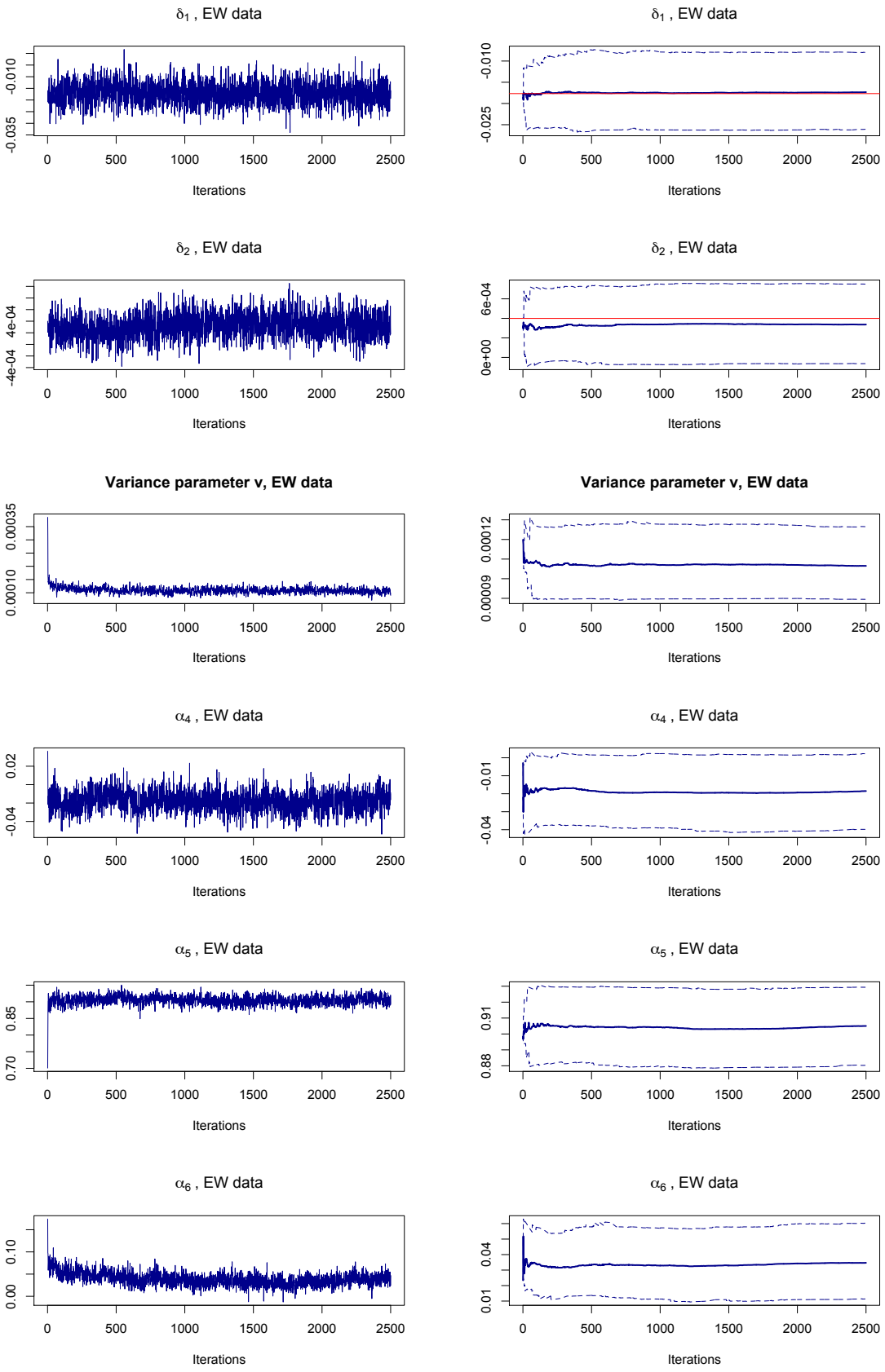

Figure 5.1: Cumulative trace-plots for the drift vector, $\boldsymbol{\delta}$, for the variance parameter, v, of matrix V_Z and for the autoregression coefficients α_4, α_5 and α_6 of the BA-CBD model, EW data. The red lines indicate the Poisson MLE's of the original CBD model.

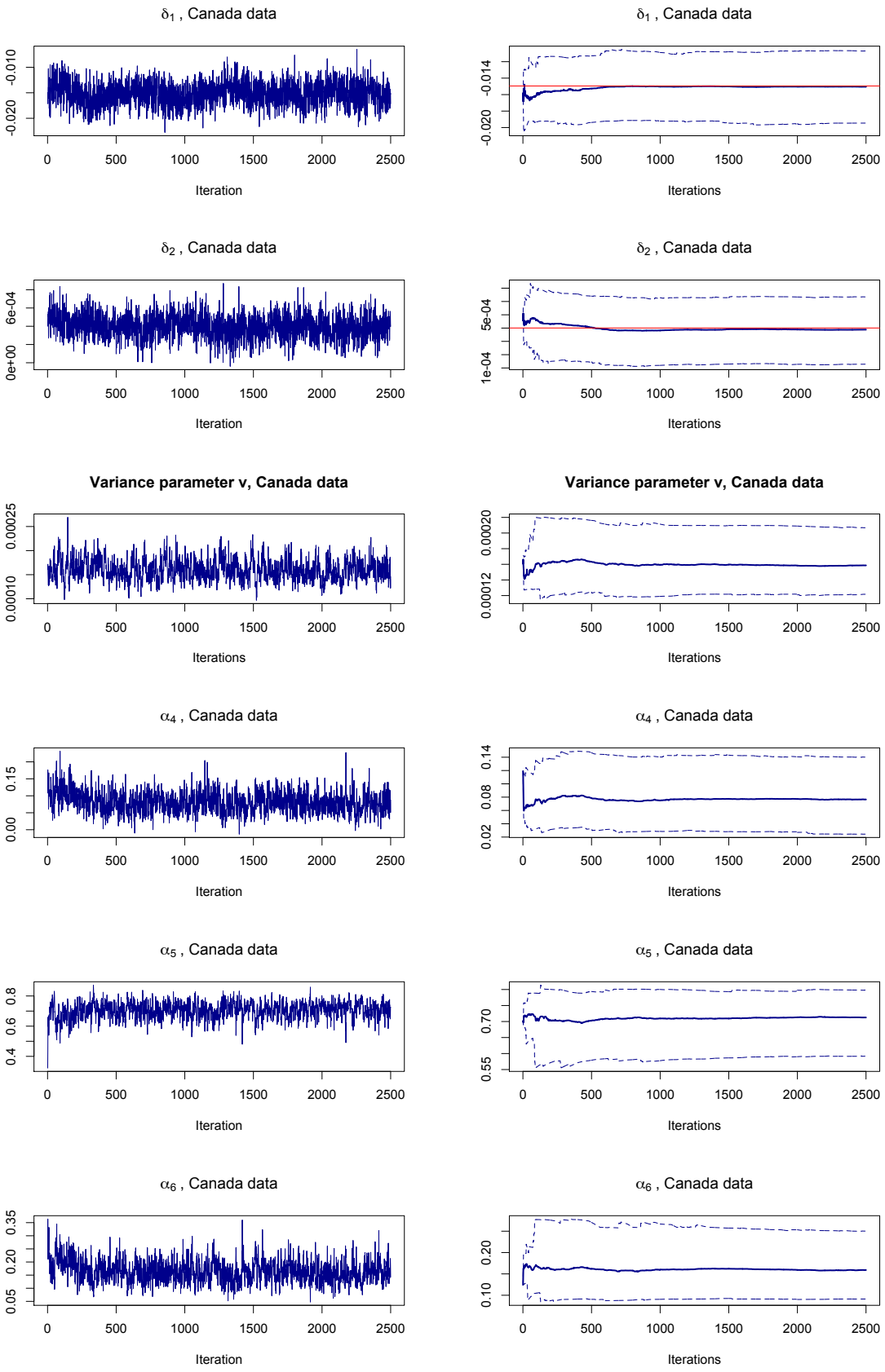

Figure 5.2: Cumulative trace-plots for the drift vector, $\boldsymbol{\delta}$, for the variance parameter, v, of matrix V_Z and for the autoregression coefficients α_4, α_5 and α_6 of the BA-CBD model, Canada data. The red lines indicate the Poisson MLE's of the original CBD model.

5.3.2 Posterior estimates

This Section presents the summaries for the posterior distribution of the BA-CBD model for the EW and Canada data-sets. The plots are based on the convergent segment of the posterior sample, according to the findings of the previous subsection. Table 5.1 summarises the posterior distributions of interest, including the results for the autoregression parameters of the matrix $\mathcal{A}$, the drift and covariance parameters of the random walk, $\boldsymbol{\delta}$ and V_ζ, and the period effects for the last observation year, $\boldsymbol{\kappa}_{2009}$, for both data-sets.

The results from the examination of the summarising tables follow closely those regarding the BA-LC model. The residuals process is again mainly driven by the cohort effect of the data, so that, for example, parameter α_5 is the dominant coefficient of the triple (α_4, α_5, α_6) for both populations. Nevertheless, α_5 is again lower in magnitude for the Canada data. However, the difference between the two data-sets is more prominent for this model compared to the BA-LC, where α_5 was again lower for Canada but much closer to that of EW. This implies a systematic lower impact of the cohort effect for the Canadian population which is supported by the plot of $\gamma_c^{(4)}$ in Figure 1.4. Since the BA-LC model yields similar values for α_5 between the two data-sets, maybe the model is not efficient in capturing the difference of the impact of the cohort effect between the two populations. Commonly with the BA-LC model, the volatilities of α_4 and α_6 are very close in the case of the EW data, but that of α_6 is almost 50% greater than the corresponding value of α_4 for the Canada population.

The Poisson MLE's of the period effects vector, $\boldsymbol{\kappa}_{2009}$, under the original CBD model are outside the 95% HPD intervals produced by the MCMC. This is in contrast to the behaviour of the period effects $\kappa_t^{(2)}$ of the BA-LC, where the resultant HPD intervals were covering sufficiently the Poisson MLE's of the original LC model. Hence, in the case of the BA-CBD model the joint estimation of the residuals states, along with the period effects, leads to a distinguishable dislocation of the posterior distributions of the latter parameters compared to the Poisson MLE's of the original CBD model. These observations become clearer when considering the time-series plots of the $\boldsymbol{\kappa}_t$ vectors in Figure 5.3. Figure 5.3 shows the mean posterior estimates of $\boldsymbol{\kappa}_t$ in black solid lines, along with the 95% credibility intervals in dashed lines. We compare these posterior distributions to the Poisson MLE's of the original CBD model and of model M7 noted in red and blue lines, respectively. The posterior distributions of $\kappa_t^{(1)}$ of the BA-CBD model show much narrower credibility bands than those of $\kappa_t^{(2)}$. This might be explained along the same lines as the lower uncertainty in the posterior distribution of $\beta_x^{(1)}$ compared to $\beta_x^{(2)}$ for the BA-LC model. Moreover, the set of $\kappa_t^{(1)}$ under the BA-CBD model is shifted in both data-sets compared to the estimates of the other two models. However, their trend is maintained for the same reason as for

the $\kappa_t^{(2)}$ of the BA-LC model. The trend similarity is also apparent and in the case of the $\kappa_t^{(2)}$.

EW data

	Mean	Std Deviation	L_{hpd}	U_{hpd}	p-value
α_1	0.98511760	0.00698467	0.96938995	0.99612870	0.855
α_2	1.33920391	0.09987630	1.13079643	1.53205380	0.459
α_3	-0.44000811	0.10487309	-0.64518556	-0.22878574	0.527
α_4	-0.01855216	0.01080725	-0.03980681	0.00220786	0.214
α_5	0.90483349	0.01280027	0.88030632	0.92936483	0.762
α_6	0.03500233	0.01253248	0.01142156	0.06024984	0.432
v	0.00010687	0.00000969	0.00008949	0.00012659	0.125
$\kappa_{2009}^{(1)}$	-3.26738045	0.01148296	-3.28775366	-3.24335655	0.672
$\kappa_{2009}^{(2)}$	0.10512280	0.00081511	0.10345469	0.10669867	0.362
δ_1	-0.01734887	0.00461631	-0.02607171	-0.00792751	0.354
δ_2	0.00033710	0.00020712	-0.00006374	0.00075009	0.011
σ_1^2	0.00101568	0.00022311	0.00066097	0.00150250	0.732
σ_2^2	0.00000206	0.00000053	0.00000122	0.00000328	0.363
σ_{12}	0.00002307	0.00000812	0.00000965	0.00004145	0.197

Canada data

	Mean	Std Deviation	L_{hpd}	U_{hpd}	p-value
α_1	0.98510157	0.00944754	0.96009454	0.99685694	0.775
α_2	1.29094694	0.21233327	0.84306169	1.69057027	0.732
α_3	-0.25589959	0.20035990	-0.61758919	0.17035309	0.444
α_4	0.07813630	0.02964133	0.02289199	0.14084821	0.182
α_5	0.70778144	0.05223144	0.59534988	0.79738001	0.354
α_6	0.16145026	0.04017092	0.08900327	0.24668422	0.202
v	0.00016064	0.00002262	0.00012110	0.00020674	0.294
$\kappa_{2009}^{(1)}$	-3.43771326	0.01499755	-3.46917440	-3.41277095	0.182
$\kappa_{2009}^{(2)}$	0.10839993	0.00107798	0.10631477	0.11041620	0.057
δ_1	-0.01516774	0.00227522	-0.01965011	-0.01064439	0.532
δ_2	0.00039307	0.00012705	0.00013420	0.00063831	0.082
σ_1^2	0.00023570	0.00005771	0.00014622	0.00036682	0.816
σ_2^2	0.00000071	0.00000027	0.00000031	0.00000133	0.225
σ_{12}	0.00000308	0.00000270	-0.00000186	0.00000882	0.406

Table 5.1: Summary of posterior distributions of the parameters of the BA-CBD model and p-value of Geweke's convergence diagnostic, EW and Canada data.

The differences observed between the $\kappa_t^{(2)}$ estimates of the Poisson CBD model and

the posterior distributions of the BA-CBD model are sufficient to cause the shift of the drift estimate noted in Figure 5.1. The latent residuals operate as with the BA-LC model so that they compensate for the structures missed by the CBD model. However, their absolute contribution to the total value of the rates is significantly lower than with the BA-LC model. Characteristically, that contribution might be as great as 4.5% for the EW data, but much lower for the Canada population.

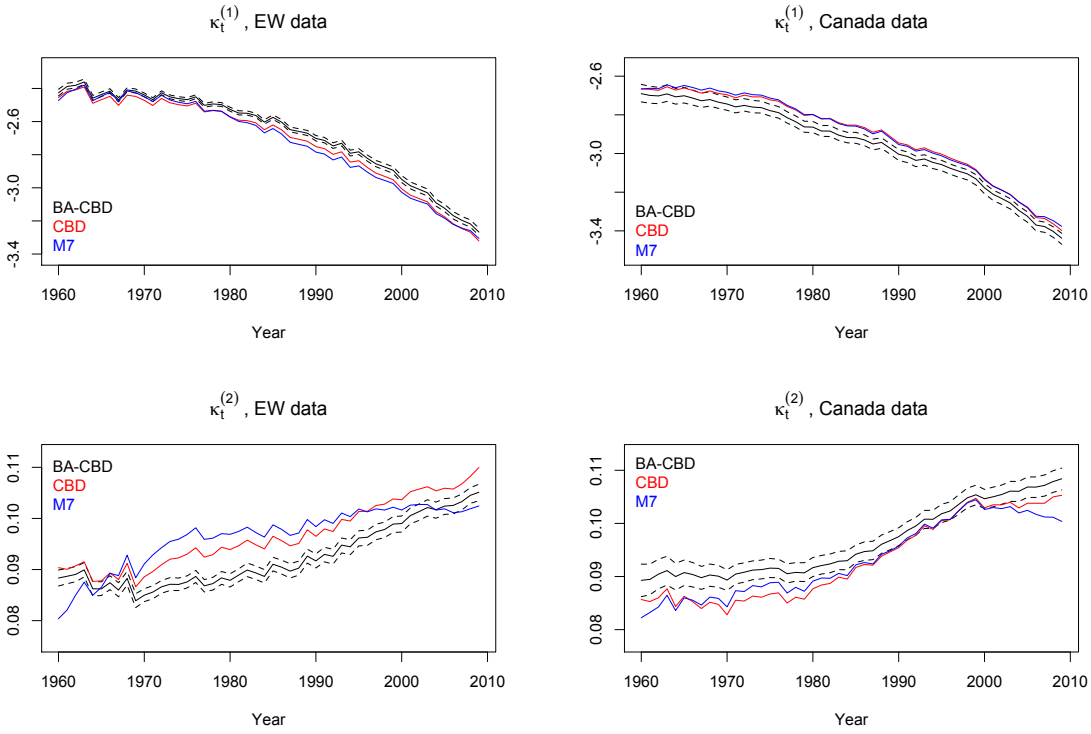

Figure 5.3: Posterior means and 95% credibility bands for $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ of BA-CBD model in black solid and dashed lines, respectively, for EW and Canada data. The original CBD and model's M7 MLE's are plotted in red and blue, respectively.

Matrix V_Z is diagonal for the BA-CBD model with the common parameter v across all ages. According to the tables of this subsection, the variability of the vectors $\boldsymbol{\mathcal{R}}_t$ is greater for the Canada data compared to EW. The stationary correlation of the $\boldsymbol{\mathcal{R}}_t$ states is shown in Figure 5.4. The plots imply that the long-term correlation between the entries of $\boldsymbol{\mathcal{R}}_t$ is stronger for the Canada data. If the underlying covariance matrices would be presented, high covariance values occur for individuals within the band of ages 60-65 for the EW data. The group of ages that develop significant covariance values for the Canada population is extended to the band 60-70.

The posterior distributions of the diagonal elements of the matrix $V_{\mathcal{R}}$ are summarised in the box-plots of Figure 5.4 for both populations. In contrast to the BA-LC model, the behaviour of the posteriors is common across the two datasets. The elements dedicated to the young ages of the data-sets are those with the higher diagonal values

and associated variability. As age increases the stationary variance component of the corresponding $\boldsymbol{\mathcal{R}}_t$ entry decreases. This observation is a result of the bivariate dynamics of the CBD benchmark model compared to the single factor of the LC model. The residuals model of the BA-LC model develops greater stationary variance for older ages due to the inability of the single factor $\kappa_t^{(2)}$ to capture the mortality patterns at old ages. On the other hand, the two period factors of the CBD model perform better in that sense and thus, the stationary variance elements are decreasing with age in the BA-CBD model.

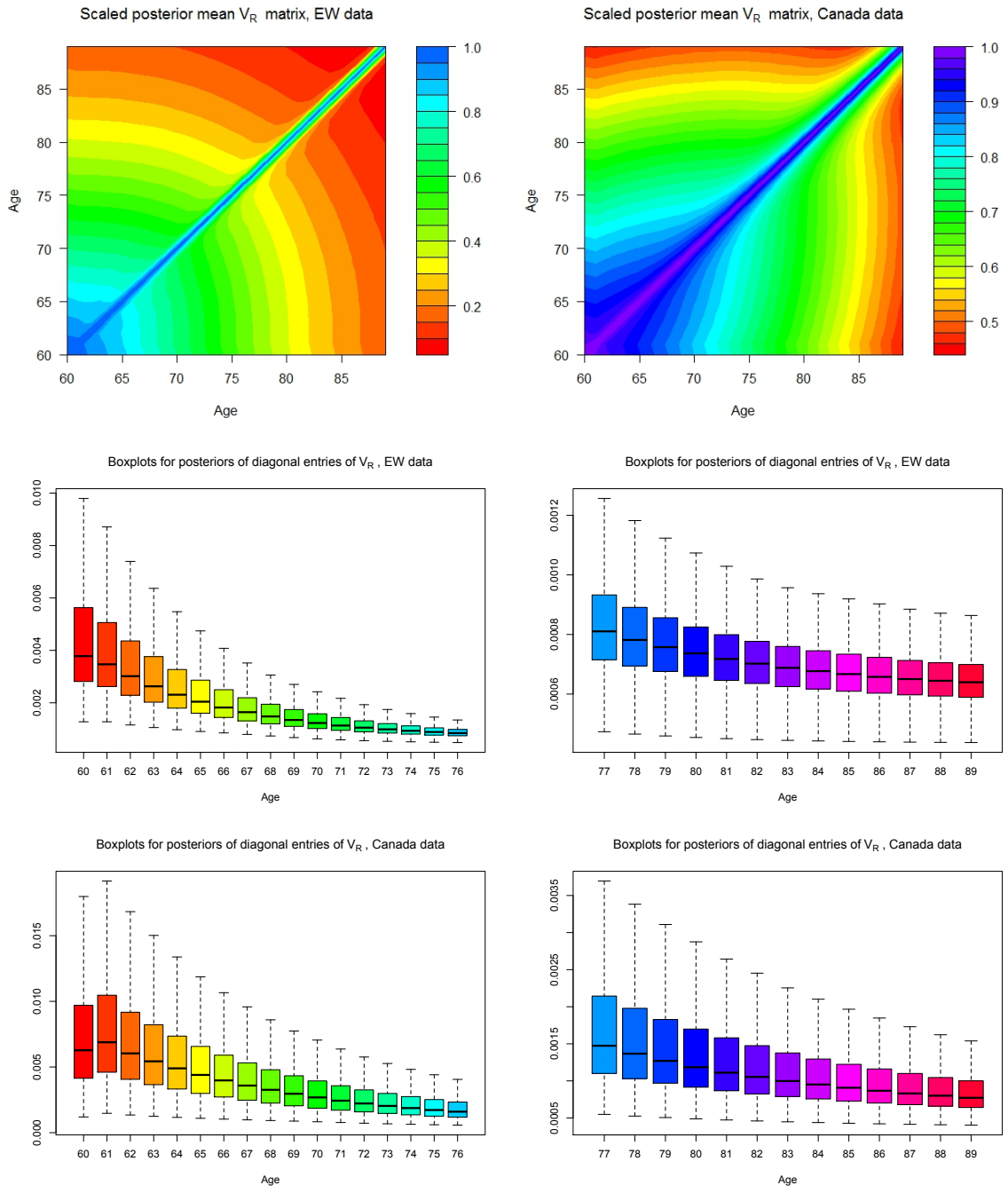

Figure 5.4: Box-plots for posteriors and posterior correlation matrices of diagonal entries of $V_{\mathcal{R}}$ matrix for EW and Canada data.

Finally, we examine the cross-correlations of the parameters within the BA-CBD model. Amongst the entries of the autoregression matrix $\mathcal{A}$, parameters α_2 and α_3 exhibit strong negative correlation for both populations. The observation is more intense in the case of the Canada data-set. Exactly the same was the corresponding result for the BA-LC model, although the relationship between α_2 and α_3 for the EW data was milder there. Consistently with the BA-LC model, the diffusion coefficients of the BA-CBD model, α_4 and α_6, appear to act inversely to the cohort-type factor of the autoregression matrix, α_5, for both data-sets. Again the observed inter-relations are stronger for the Canada population. Interestingly, the joint posterior of parameters α_4 and α_6 shows positive dependence in the case of the EW data, but the opposite is true for the Canada population. Lastly, the parameter, v, of the covariance matrix V_Z appears to be clearly correlated with the triple α_4, α_5 and α_6 of the autoregressive matrix, $\mathcal{A}$. For both populations v is positively related to α_6 and negatively related to the cohort-type coefficient α_5. In the case of the EW population any relationship almost disappears when considering the joint density of α_4 and v, in contrast to the Canada data-set.

It is notable that the drifts, $\boldsymbol{\delta}$, appear *a posteriori* positively correlated, more prominently for the EW data, although this is not the case in an MCMC estimation of the CBD model alone, as commented in Section 2.4.1. The same scale of the plots in Figure 5.5 also aids in identifying the increased posterior uncertainty in the estimates of $\boldsymbol{\delta}$ for the EW population compared to the Canada data. However, for both populations the *IW* sampling of the covariance matrix yields reasonable correlation between its entries. Again in the case of the EW data-set the phenomenon is more intense.

Similar plots have been produced to examine relationships amongst the latent period and residuals states, as well as dependencies of those states to the parameters of their underlying models. The latent effects $\boldsymbol{\kappa}_t$ and $\boldsymbol{\mathcal{R}}_t$ for fixed year t are verified to act independently amongst them and no other important cross-relationships appear to exist in that sense, beyond signs of negative correlation between the period effects vectors for fixed time, t.

In this section, we have analysed the results of the application of the BA-CBD model to the EW and Canada data-sets. After assessing the convergence of the implemented algorithms, the resultant posterior distributions were summarised and their estimates, where applicable, were compared to Poisson MLEs of the original CBD model. Many of the results follow as in the case of the BA-LC model. For example, the same dependence structure was observed amongst the elements of the autoregressive matrix $\mathcal{A}$. Similarly, the posterior distributions for the latent state estimates for the period effects are significantly distorted, but the main trend of the time-series is maintained.

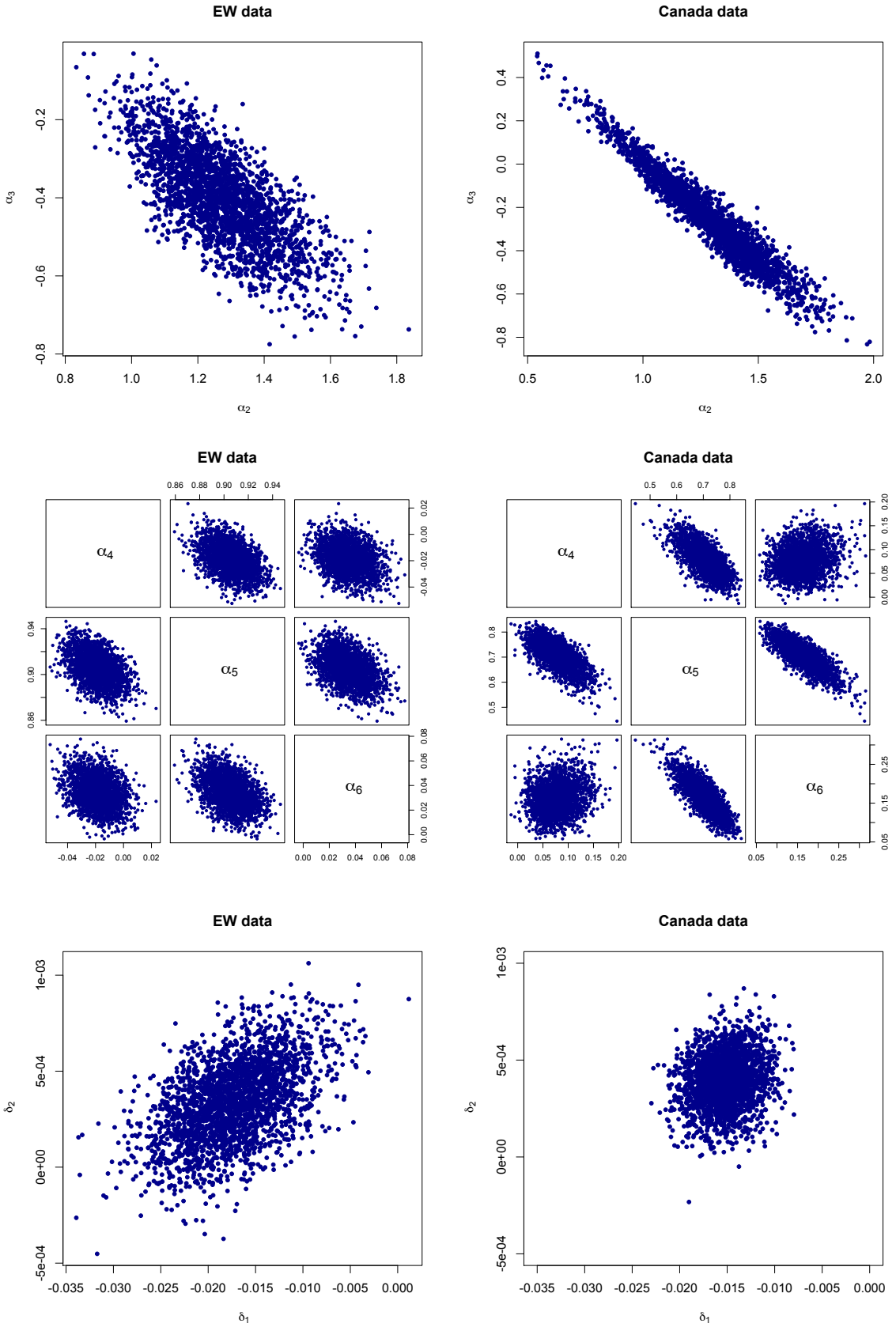

Figure 5.5: Bivariate scatterplots for posterior distributions of the drift vectors, $\boldsymbol{\delta}$, and the autoregression parameters, α_2 and α_3, and α_4, α_5 and α_6, for EW and Canada data.

5.4 Summary

This Chapter developed the BA-CBD model along the lines of the BA-LC model. The common use of the Poisson model for deaths aids in comparing the performance of the two models under the same source of idiosyncratic risk in Chapter 6.

In this case the long-term mortality dynamics are driven by the period factors of the CBD model. The existence of this pair of stochastic processes allows a more parsimonious structure for the residuals model and results in a more stable estimation algorithm. The joint prior distribution of the full model is only focused on specifying information about parameter v on the diagonal of the covariance matrix V_Z, and setting the appropriate supports for the autoregression parameters of matrix A. The conjugate property of $\boldsymbol{\delta}$ and V_ζ allows for Gibbs steps in the estimation scheme, while MH sampling is employed for the latent states $\boldsymbol{\kappa}_t$, $\boldsymbol{\mathcal{R}}_t$ and the elements of matrix $\mathcal{A}$. Finally, an IG approximate proposal is used for the variance parameter v. The resultant hybrid MCMC algorithm is a combination of **Algorithm 3** developed in Section 2.4.1 and of the residuals model sampling part of **Algorithm 4** given in Section 4.2.2.

The historical long-term dynamics of the BA-CBD model, and the parameters driving them, are very similar to those of the original CBD model for both examined data-sets. The slope parameters $\kappa_t^{(2)}$ appear to have the highest posterior uncertainty, whereas the interpretation of the autoregression is kept consistent to its definition. The posterior correlation within the elements of the autoregression matrix which was noted in Chapter 4 is also observed in this case and much more intensively than for the BA-LC model. Again, the Canada population develops that effect more prominently. Finally, it is worth noting the greater uncertainty in the joint posterior of the drift $\boldsymbol{\delta}$ for the EW data, and that this is only being developed after the inclusion of the residual states within the hierarchical model. That form of randomness, along with the greater in absolute value estimates for the parameters of V_ζ, will generally produce wider confidence bands around the forecasted rates for the EW population in the following Chapter.

Chapter 6

Comparison of proposed models

In this Chapter the BA-LC and BA-CBD models, commonly named BA-models henceforth, are subject to qualitative and quantitative comparison after their calibration to the EW and Canada data. The goodness-of-fit is examined through the fitted rates for the final observation year and the matrices of standardised residuals. The dynamics of the BA-models are assessed according to their level of persistence. The Deviance Information Criterion (DIC) is employed as a means of quantitative comparison within the hierarchical modelling framework. We compare the projected mortality rates and confidence levels between the two BA-models, and for each BA-form within its respective family. Similarly to Section 1.4.3, we examine the produced survivor index distributions and related quantities. Given the realised rates in year 2010 for EW, the predictive performance of all stochastic mortality models is compared. The forecasted mortality improvements and their uncertainty levels are considered, first for the first projection year, and second for the full projection period. Finally, the robustness properties of the BA-models are analysed by fitting the models to a subset of the EW population. The natural interpretation of the local mortality dynamics and the similar behaviour of the long-term dynamics are then verified. Also, we proceed to in-sample forecasts and conclude which of all the stochastic mortality models of the Thesis has had the best predictive performance for the decade 2000-'09.

6.1 Goodness-of-fit

The quality of fit of the proposed mortality models is examined by the fitted rates they produce and through the properties of the posterior distribution of their standardised residuals matrices. The fitted rates are also compared to those produced by the mortality models of Chapter 1. The results are investigated according to the benchmark family each model belongs to. The posterior distributions of the standardised residuals matrices are examined graphically, and are also assessed using the statistical tests of Chapter 3.

6.1.1 Fitted rates

The fitted rates of the BA-models are obtained by considering the thinned full convergent posterior sample, as determined by the relevant sections for each extension. For each iteration the algorithm provides a matrix of fitted rates, and that sample may be summarised to obtain the posterior distribution of the fitted rates.

The BA-models are examined separately for each family. Figure 6.1 shows the observed rates across all ages for the jump-off year 2009. The output of the CBD-type models has been transformed to central death rates using relationship (1.1.7) to facilitate comparison. In each graph we include the results of the rest members of the family. So the top panels of Figure 6.1 include the observed and fitted rates of the LC, APC and BA-LC models. Similarly, the bottom panels illustrate the quality of fit of the CBD, M7 and BA-CBD models. The distributions of the fitted rates of the BA-models are summarised via their posterior medians, while their minimal spread is noteworthy. Characteristically, any attempt to graphically illustrate any of their sensible percentiles fails since a typical value of the coefficient of variation for the posterior of the fitted rates for any age is around 0.5% and the resulting plots are not comprehensive. The examination of the final observation year aids in getting a better insight into the way the stochastic mortality models initiate the forecasting procedure. The estimates of the BA-models almost match the observed rates and capture the majority of the mortality jumps between ages. The results for the examined year show that the BA-models fit the data better than the rest stochastic mortality models. The simple LC and CBD models systematically underestimate the observed mortality rates for young ages of the population. The CBD-type of models appear to perform slightly better in terms of goodness-of-fit. In particular, the APC model produces rates that deviate significantly from the observation especially for the first five ages of the data-set. However, if the same investigation would be carried out for a different calendar year, it could be seen that there are instances where the the stochastic mortality models which incorporate cohort parameters might perform very similarly

to the BA-models. The BA-models or the direct cohort-extended versions of the basic mortality models do not assume smoothness at any dimension of the mortality table. That feature seems essential for an accurate fit to the observed rates, and therefore a good starting point for the forecasts, since in all cases the benchmark mortality model of the respective family yields the less effective fit to the observations.

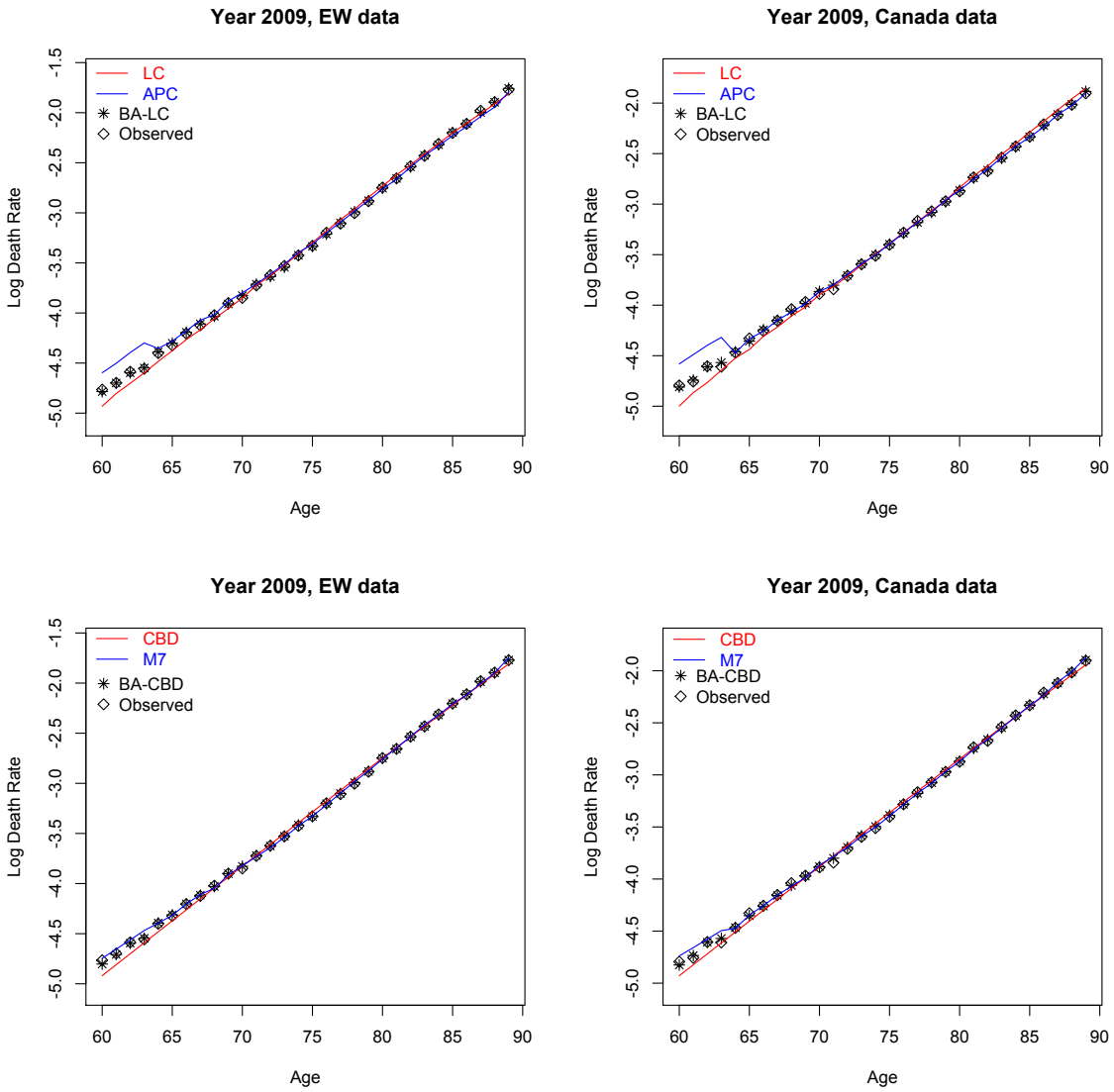

Figure 6.1: Observed log-death-rates and fitted estimates of the LC, CBD, APC, M7, BA-CBD and BA-LC models for ages 60-89 in year 2009, EW and Canada.

Given the fitted rates for the final observation year and the forecasts for year 2010, it is possible to derive the distribution of the improvement rates for the first year of the projection. A detailed examination of the improvement rates over the whole projection period is conducted in Section 6.5. However, the improvements for the first projection year therein appear to be extreme and distort a clear view of the magnitudes in the respective contour plots of Figures 6.17-6.22. Hence, we examine that first year as part of this Section in isolation. Figure 6.2 shows the graphs of the mean improvement

rates and their associated 95% confidence bands for all stochastic mortality models and both data-sets. The plots show that the LC and CBD models predict extremely high mean improvement rates for early ages of both data-sets. Model M7 behaves distinctly in the case of EW data, since for the first few ages the mean improvement rates vary from -20% to -10%, in contrast to the results of the rest of the stochastic mortality models. Finally, the improvements are more volatile for the EW data, which is mainly due to the smoother stochastic process underlying the Canadian population.

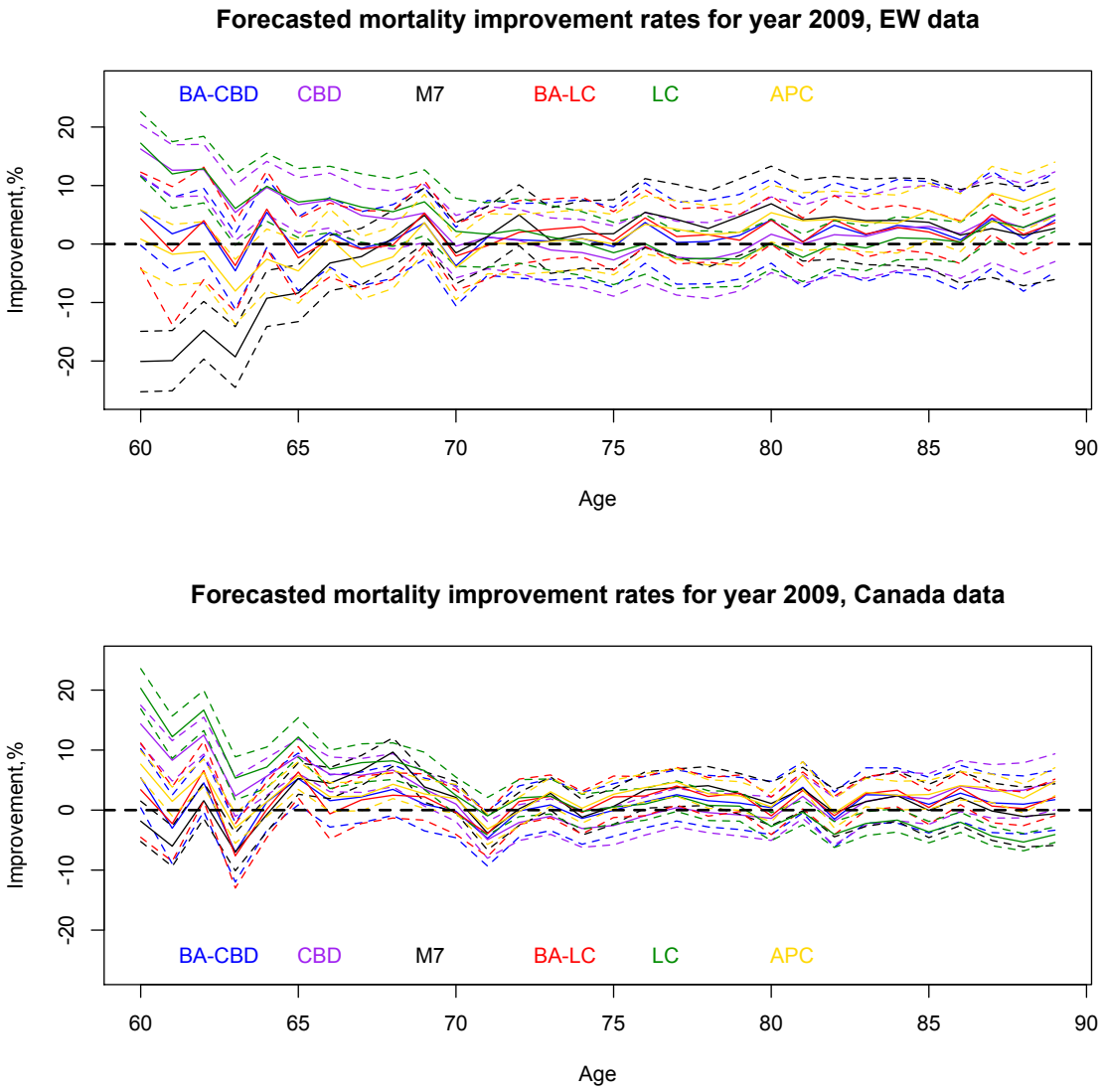

Figure 6.2: Projected mean mortality improvement rates and associated 95% confidence bands for ages 60-89 in year 2009, EW and Canada data.

6.1.2 Standardised residuals

The posterior distributions of the standardised residuals of deaths are obtained by constructing the fitted rates for each parameter vector of the convergent sample and using formula (1.4.1). The standardised residuals distribution for each age-year cell of the data under the BA-models are summarised by their posterior medians which are plotted in Figure 6.3 in red and black squares for positive and negative values, respectively. By comparing the plots to those presented in Figure 1.5 of Section 1.4.1, it is evident that the BA-models model performs superiorly to the basic model of each family. The slight evidence of clustering in the graphs of the cohort enhanced models of each family seems to mitigate in the figures of the BA-models, where no identifiable pattern may be observed.

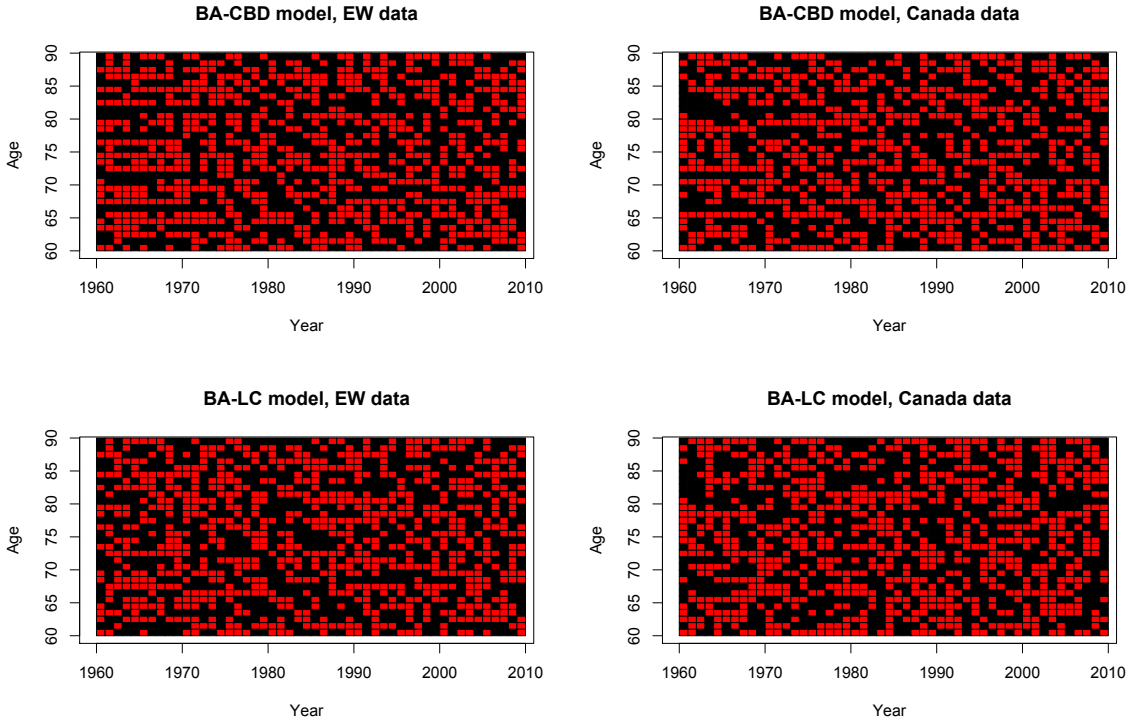

Figure 6.3: Graphs of posterior means of standardised residuals for the BA-CBD and BA-LC models for EW and Canada data, indicated in red and black squares for positive and negative values, respectively.

We also apply the statistical tests employed in Section 3.2.2 at a 1% confidence level for each standardised residuals matrix of the MCMC sample. As already mentioned, each constructed residuals matrix is required to consist of approximately *i.i.d.* standard normal variates both when one follows a certain age from one year to another and when death rates are compared over the range of ages for any fixed year. Table 6.1 summarises the mean rejection rates of the sample series across the age and year dimensions regarding the hypotheses of independence (LB test), normality (JB test),

zero-mean (t-test) and unitary variance (χ^2-test), over the posterior distribution of the standardised residuals matrices. The results are equivalent between the two populations and indicate that the examined distribution satisfies the underlying hypotheses. For both extensions, the independent, zero-mean normality of the examined time series is well supported by the posterior distribution of the resulting matrices. The results of the BA-CBD model for the EW data-set show increased rejection rates for the unitary variance assumption for the age series. The values which reject the test always fall below 1, which implies a tight fit of the model. However, for each logit-rate modelled, an auxiliary variable $\mathcal{R}(x,t)$, has been added to the model. The addition of those latent states allows the determination of the parameters of the model for $\boldsymbol{\mathcal{R}}_t$. The tight fit of the model allows initiating the projections almost from the observed rates.

BA-LC model

	EW Data		Canada Data	
	Age Series	Year Series	Age Series	Year Series
Runs test	4.52%	4.01%	4.06%	4.19%
JB test	6.47%	4.84%	3.99 %	2.99%
t-test	32.84%	5.23%	18.41%	4.78%
χ^2-test	7.21%	13.08%	7.79%	9.05%

BA-CBD model

	EW Data		Canada Data	
	Age Series	Year Series	Age Series	Year Series
Runs test	5.98%	3.12%	5.09%	4.63%
JB test	4.95%	3.36%	3.89%	3.55%
t-test	4.52%	4.88%	4.11%	4.73%
χ^2-test	5.14%	5.18%	5.14%	5.01%

Table 6.1: Mean rejection rates of statistical tests, at a 5% confidence level, over the posterior distribution of the standardised residuals matrix of the BA-models for EW and Canada data.

Towards investigating further the underlying hypotheses regarding the residuals process, $\boldsymbol{\mathcal{R}}_t$, of the BA-models we also calculate the latent residuals of the VAR model. That is, for each iteration j of the convergent sample, we calculate:

$$\boldsymbol{Z}_t^{(j)} = \boldsymbol{\mathcal{R}}_t^{(j)} - \mathcal{A}^{(j)}\boldsymbol{\mathcal{R}}_{t-1}^{(j)}, \text{for } t = 2, \ldots, n.$$

According to the assumptions of the BA-models, $\boldsymbol{Z}_t^{(j)}$ are predicted to be *i.i.d.* series of zero-mean, normally distributed, m-dimensional vectors across time t, for each iteration j. We apply the multivariate version of the Shapiro-Wilks normality test, also employed in Section 3.3.2, to each $\boldsymbol{Z}_t^{(j)}$ matrix of the posterior sample at a 5% confidence level for both models. The sample of calculated p-values supports overwhelmingly the normality of the $\boldsymbol{\mathcal{R}}_t$ series for both populations. Subsequently, the Hoteling test is applied to investigate the zero-mean hypothesis for the series of $\boldsymbol{Z}_t$ vectors across their posterior distribution. The results for both data-sets and models strongly back that assumption too.

The BA-models provide tight fit, allowing the initiation of the forecasts almost from the observed rates. In some cases the suggested models might perform similarly to cohort-enhanced versions of the original families. However, across the whole set of ages and range of years of each data-set, the BA-models almost match the historical rates. Hence, the models take into account a great amount of information about the mortality characteristics of each population. That will become more evident in the comparative out and in-sample projections, where it will be seen how this population information is incorporated until the cohorts exit the boundaries of the projected mortality table.

6.2 Model dynamics

In this part we examine the posterior distributions of the parameters of the BA-models that determine the future dynamics. The local mortality dynamics are driven by the residuals model; hence, we are interested in the form of the autoregression matrix, $\mathcal{A}$, and the covariance structure of the disturbances of $\boldsymbol{\mathcal{R}}_t$. The latter is summarised via the marginal distribution of the single parameter of the diagonal v in the case of the BA-CBD model, and via the element-wise median of the posterior distribution of the V_Z matrix for the BA-LC model. The long term mortality dynamics are controlled by the parameters of the random walk models in both cases. Thus, for the BA-LC model we are interested in the constant drift, δ, and the constant variance, σ_κ^2, whereas in the case of the BA-CBD model we care for the drift vector, $\boldsymbol{\delta}$, and the parameters of the 2×2 covariance matrix, V_ζ.

6.2.1 Long-term dynamics

Figures 6.4 and 6.5 include the marginal posterior density plots of the parameters which control the long-term dynamics of the BA-CBD and BA-LC models in blue and red, respectively. The results are shown in solid for EW and dashed for Canada.

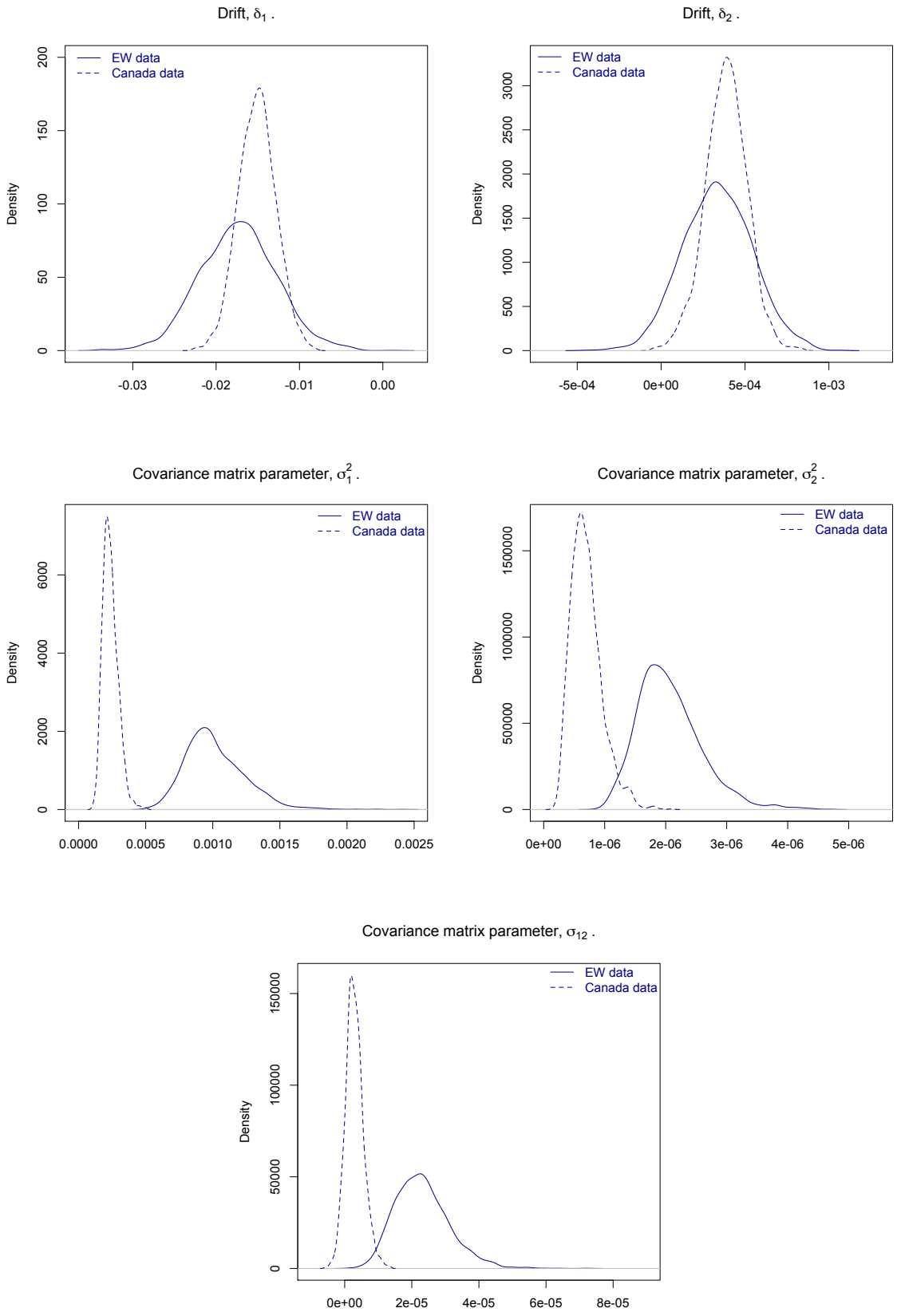

Figure 6.4: Marginal posterior densities of the long-term dynamics parameters, $\boldsymbol{\delta}$ and V_ζ, of the BA-CBD model for EW and Canada data.

The distinctive behaviour of the marginal posterior distributions for the parameters of the long-term dynamics of the BA-models between the two populations results in qualitative differences in the relevant projections. The marginal distributions of the drift vector components, $\boldsymbol{\delta}$, of the BA-CBD model are concentrated around similar values for both populations. However, the posterior spread for the drift vector parameters of the EW data are much greater than that of the Canada population. The identical observation is true for the drift, δ, of the BA-LC model between the two data-sets. Although the projections for the two populations under the BA-models would incorporate similar drift values on average, the increased uncertainties of their posterior distributions for the EW data contribute to the widths of the respective confidence bands. Similarly to the drifts' behaviour, the posterior distributions of the co/variance structure for the random walk components of the BA-models develop reduced posterior spread for the Canada population. In accordance to the MLE results of Chapter 1, their marginal posteriors for the Canada data-set are also located in much lower values compared to the EW data. Characteristically, the posteriors for σ_1^2 of the BA-CBD model for the two populations hardly overlap. Therefore, beyond the smaller co/variance in the long-term dynamics of the Canada data, we also have lower posterior variability for that co/variance term(s). That difference would not be apparent and incorporated if parameter uncertainty was not imposed through the usage of the relative posterior distributions.

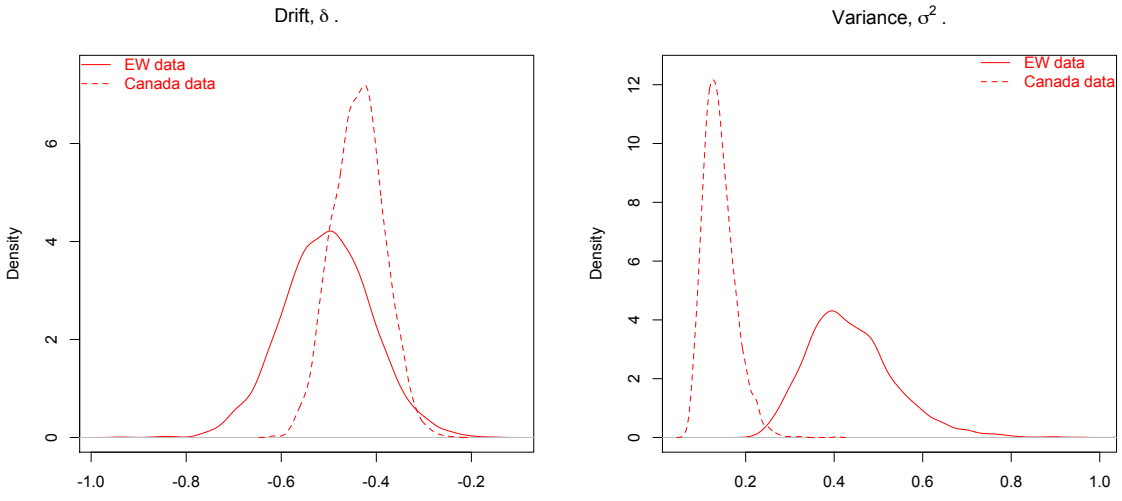

Figure 6.5: Marginal posterior densities of the long-term dynamics parameters, δ and σ_κ^2, of the BA-LC model for EW and Canada data.

6.2.2 Local dynamics

The short-term dynamics of the BA-models are driven by the residuals process. Figure 6.6 shows the marginal posterior densities for the parameters of the autoregression matrix $\mathcal{A}$ for both BA-models and populations.

The elements of the autoregression matrix determine the short-term mean for the residuals model. Parameter α_1 serves for the autoregressive relationship of the first entrants of the data-set. In the case of the BA-CBD model the marginal posteriors are very close to the bound of 1 and look very similar for both data-sets. In contrast, for the BA-LC the posterior sample is constituted by lower values but with much greater spread. Particularly for the EW population, α_1 develops a strongly skewed distribution with noticeable left tail. Parameter α_2 is *a posteriori* located around very similar values between both populations for the BA-CBD model, though the tails of the distribution in the case of the Canada data are evidently more heavy than those for the EW data. On the other hand, the cohort-type coefficient α_2 behaves quite distinctively in the case of the BA-LC model, with the posterior distribution of the Canada population being located much lower. The shape of the posterior distributions for the diffusion coefficient α_3 is very similar between the two BA-models, but with the values of the BA-CBD for the EW data being much smaller.

The parameters examined so far are only used to produce the short-term mean contribution of the residuals model for the first two ages of the data-set. The expected residuals values for the rest of the ages are constructed from the triple $(\alpha_4, \alpha_5, \alpha_6)$. Hence, that triple is responsible for the mean contribution of the residuals model for the majority of the ages in the data-set. All three parameters, α_4, α_5 and α_6, have marginal posterior distributions with greater spread for the Canada data under both BA-models. The posterior range for the diffusion coefficient α_4 is comparable between the two BA-models for both populations. In contrast, the posterior for parameter α_5 is very different between the two BA-models. For both data-sets, the results of the BA-CBD model are located much lower than those of the BA-LC model and hardly overlap. The identical comments apply for the diffusion coefficient α_6 between the two data-sets for the two BA-models. Characteristically, in both cases the marginal posteriors of the BA-CBD model for the EW data overlap with the corresponding densities of the BA-LC model for the Canada data.

Finally, we examine the short-term covariance structure of the disturbances of the residuals models. Figure 6.7 summarises the element-wise posterior mean of matrix V_Z for the BA-LC model and the marginal posterior density for parameter v of the diagonal of matrix V_Z for the BA-CBD model. In terms of the BA-CBD model, the distribution of parameter v is located significantly lower for the EW data compared

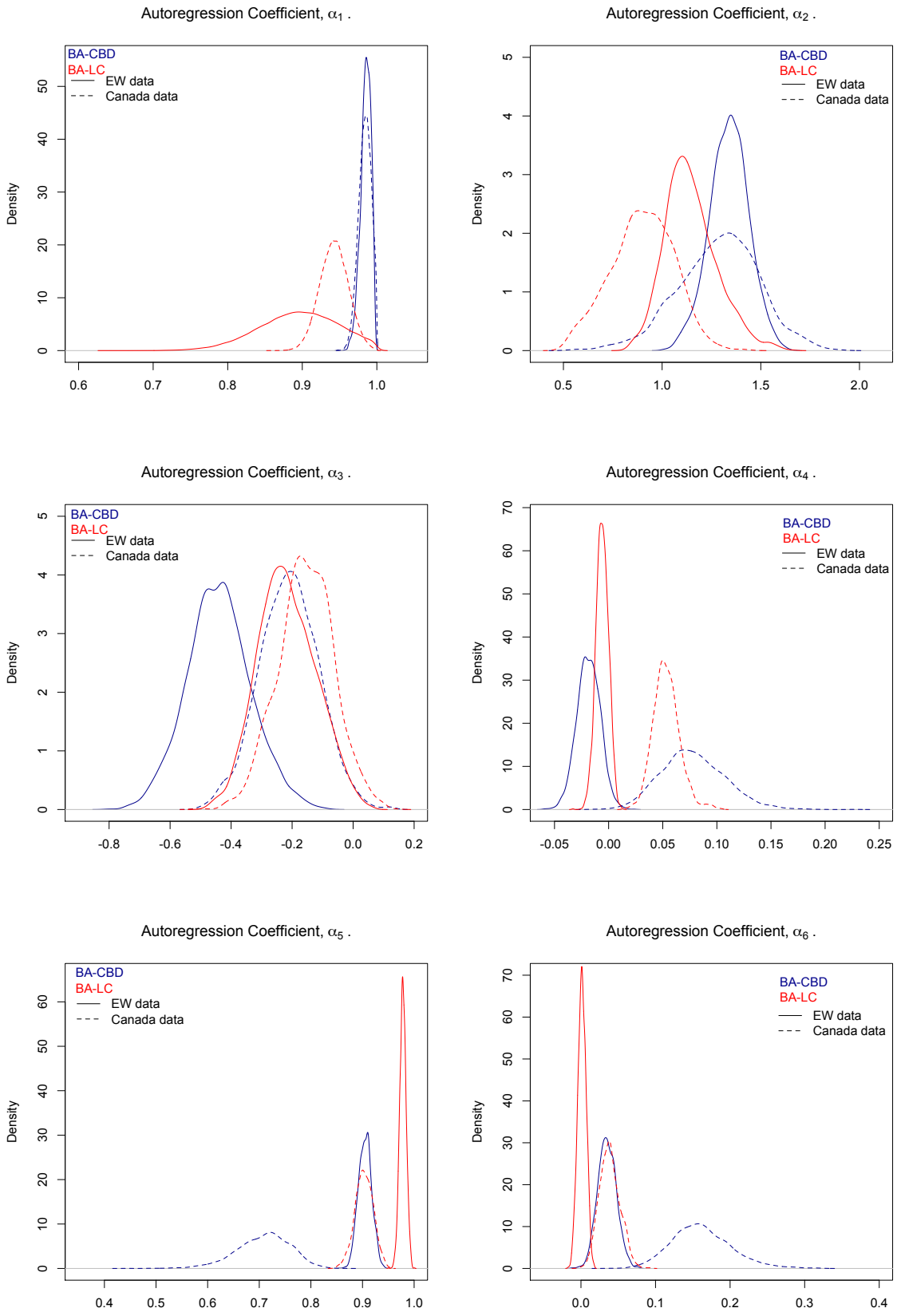

Figure 6.6: Marginal posterior densities for the parameters of the autoregression matrix, $\mathcal{A}$, of the BA-models for EW and Canada data.

to the Canada population, while the associated uncertainty is also greater for the latter case. Regarding the BA-LC model, matrix V_Z develops a clear structure in the case of the EW population. The residuals terms for higher ages within the data-set are more strongly correlated and with a greater variance component than for young ages. The negative correlation between the residuals terms of young and old ages is apparent in both graphs, although the magnitude of the dependence is much lower in the case of the Canada population. In retrospect, the plot for the mean posterior covariance matrix V_Z of the BA-LC model for the Canada data-set would imply that a diagonal structure as in the case of the BA-CBD model could suffice for the specification of the model. However, the clearly structured positive and negative covariance segments within the matrix, albeit small in absolute magnitude, do require the more complex specification so that the chain is able to reach an admissible posterior distribution.

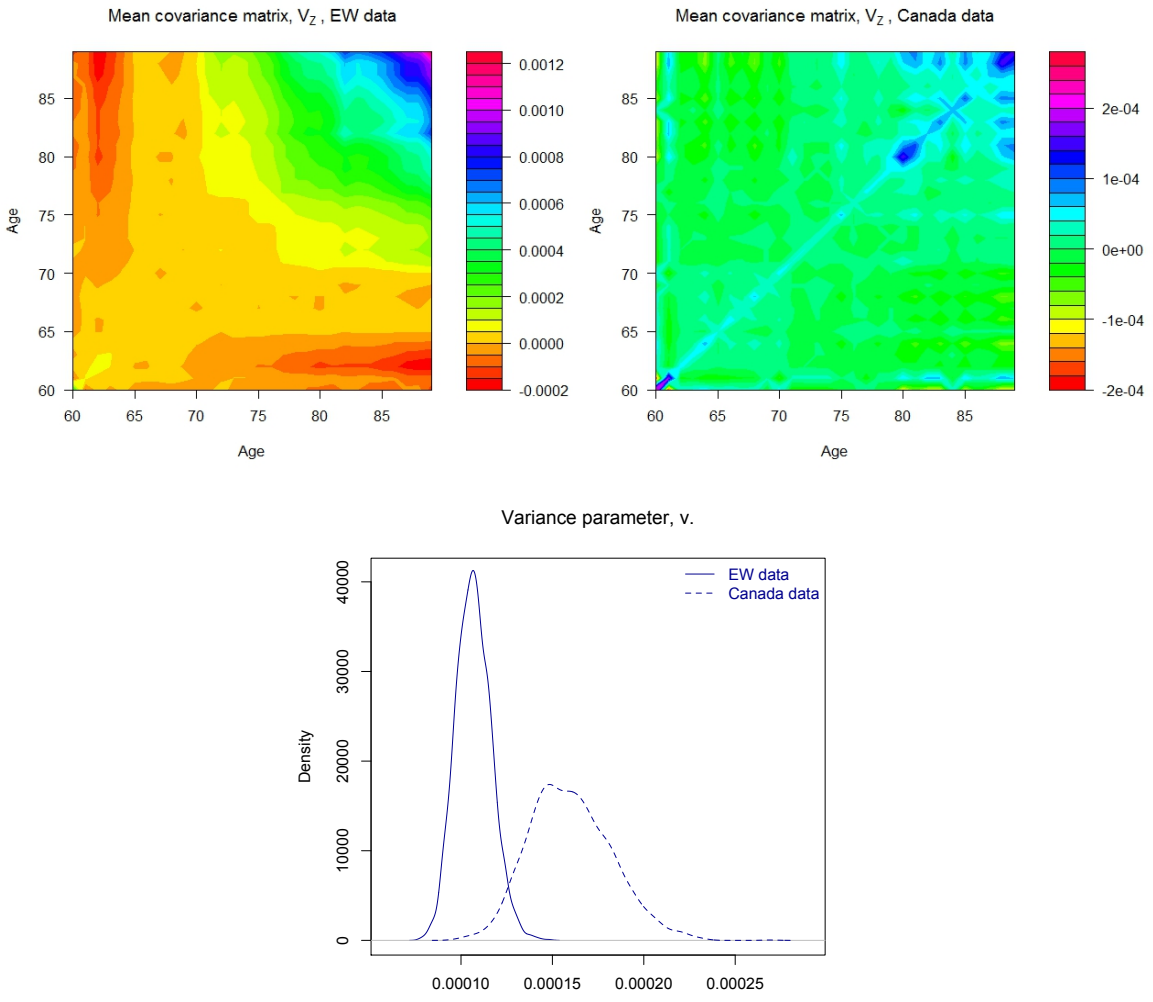

Figure 6.7: Marginal posterior summaries for the covariance structures of the BA-models for EW and Canada data.

The dynamics of the proposed extensions have been compared by distinguishing their qualitative impact. The persistent, in the long-term, parameters of the random walk models generally develop marginal posterior distributions with greater uncertainty for the EW data. Astonishingly, the opposite is true for the parameters of the local mortality dynamics as these are expressed through the marginal posteriors of the autoregression elements of matrix $\mathcal{A}$ and the corresponding covariance structure for each examined case.

6.3 Deviance information criterion

In this section we compare the BA-models based on the deviance information criterion (DIC), which follows closely the lines of the BIC used in Chapter 1. In general, parametric statistical modelling of data $\boldsymbol{y}$ requires the specification of a relevant probability model, $f(\boldsymbol{y}|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of parameters. Associated with the specified model, the corresponding deviance is the quantity:

$$\mathcal{D}(\boldsymbol{\theta}) = 2\log\left(f\left(\boldsymbol{y}|\boldsymbol{\theta}\right)\right) - 2\log\left(h(\boldsymbol{y})\right),$$

where $h(\boldsymbol{y})$ is a function of the data alone.

The deviance statistic is the typical measure of fit under the frequentistic modelling framework. Supplementing this measure of fit with a measure of complexity, usually the degrees of freedom of the model, *i.e.* the number of unconstrained parameters, or some other quantity based on it, leads to information criteria, such as the BIC which was used in Chapter 1, that compare different parametric models in order to choose the most appropriate one. However, in complex hierarchical models, such as those developed in Chapters 4 and 5, the number of parameters outnumber observations and thus, these methods are thus inadequate. Spiegelhalter et al. (2002) propose a complexity measure for such high-dimensional parameter spaces by defining the effective number of parameters for a model, $p_{\mathcal{D}}$. Their definition of $p_{\mathcal{D}}$ is:

$$p_{\mathcal{D}} = \overline{\mathcal{D}(\boldsymbol{\theta})} - \mathcal{D}(\widehat{\boldsymbol{\theta}}),$$

where $\overline{\mathcal{D}(\boldsymbol{\theta})} = \overline{\mathcal{D}}$ is the posterior mean deviance and $\widehat{\boldsymbol{\theta}}$ is some posterior estimate of the parameter vector. Originally, the criterion was developed based on the posterior mean, but it can be extended well to the relevant mode or median of $\boldsymbol{\theta}$.

When comparing competitive models it is sensible to choose $h(\boldsymbol{y}) = 1$ in the deviance definition, since the data need to be the same for all implementations and the term becomes irrelevant (Celeux et al., 2006). Further, the mean posterior deviance, $\overline{\mathcal{D}}$,

formally defined as:

$$\overline{\mathcal{D}} = 2\mathbb{E}_{\boldsymbol{\theta}}\left[\log\left(f(\boldsymbol{y}|\boldsymbol{\theta})\right)|\boldsymbol{y}\right],$$

is used as a Bayesian measure of adequacy, and augmented with the aforementioned measure of complexity, $p_{\mathcal{D}}$, forms a deviance information criterion as follows:

$$DIC = \overline{\mathcal{D}} + p_{\mathcal{D}}.$$

The DIC might now be expressed directly in terms of the likelihood of the model as:

$$DIC = 4\mathbb{E}_{\boldsymbol{\theta}}\left[\log\left(f(\boldsymbol{y}|\boldsymbol{\theta})\right)|\boldsymbol{y}\right] - 2\log\left(f\left(\boldsymbol{y}|\widehat{\boldsymbol{\theta}}\right)\right).$$

The argumentation might be extended in models with several layers of hierarchy. For a model whose parameters are further parameterised, Celeux et al. (2006) develop several variations of the DIC so that it coherently incorporates latent variables models. Regarding the mortality models suggested in this Thesis, the parameter vector $\boldsymbol{\theta}$ might be decomposed into two sub-vectors $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$. Vector $\boldsymbol{\phi}$ includes the parameter sets that construct the underlying rates, so that $\boldsymbol{\phi} = \left\{\beta_x^{(1)}, \beta_x^{(2)}, \kappa_t^{(2)}, \mathcal{R}(x,t)\right\}$ for the BA-LC model and $\boldsymbol{\phi} = \left\{\kappa_t^{(1)}, \kappa_t^{(2)}, \mathcal{R}(x,t)\right\}$ for the BA-CBD model. Vector $\boldsymbol{\psi}$ comprises the parameters of the stochastic models for the time-series components BA-models. Hence, $\boldsymbol{\psi} = \left\{\delta, \sigma_\kappa^2, \mathcal{A}, V_Z\right\}$ for the BA-LC model and $\boldsymbol{\psi} = \left\{\boldsymbol{\delta}, V_\zeta, \mathcal{A}, v\right\}$ for the BA-CBD model.

Two of the various proposed forms of DIC introduced by Celeux et al. (2006) are used here. Firstly, the *observed* DIC, based on the observed (Poisson) likelihood of the model, where $\widehat{\boldsymbol{\phi}}$ is taken as the posterior mean of $\boldsymbol{\phi}$, so that:

$$DIC_O = 4\mathbb{E}_{\boldsymbol{\phi}}\left[\log\left(f(\boldsymbol{y}|\boldsymbol{\phi})\right)|\boldsymbol{y}\right] - 2\log\left(f\left(\boldsymbol{y}|\widehat{\boldsymbol{\phi}}\right)\right),$$

and secondly, the *complete* DIC, based on the full likelihood of the model, where $\boldsymbol{\psi}$ is again estimated by mean of the MCMC sample, so that:

$$DIC_C = 4\mathbb{E}_{\boldsymbol{\phi},\boldsymbol{\psi}}\left[\log\left(f(\boldsymbol{y}|\boldsymbol{\phi},\boldsymbol{\psi})\right)|\boldsymbol{y}\right] - 2\log\left(f\left(\boldsymbol{y}|\widehat{\boldsymbol{\phi}},\widehat{\boldsymbol{\psi}}\right)\right).$$

	EW data		Canada data	
	DIC_O	DIC_C	DIC_O	DIC_C
BA-CBD	-19,397.9	-7,607.0	-16,572.9	-5,333.5
BA-LC	-31,554.3	-16,235.4	-22,594.9	-6,828.5

Table 6.2: Deviance Information Criteria for the Proposed Models

In the form described here, the greater the value of the DIC, the better the performance of the model. Both criteria rank higher the augmented form of the CBD model overwhelmingly.

6.4 Forecasted rates

The forecasted rates of the BA-models are obtained by considering the full convergent posterior sample. Out of that posterior sample a large number of iterations are picked at random, and for each chosen iteration every model is projected. In order to reduce the sampling error as much as possible, we randomly choose 2,500 iterations out of the 15,000 or 20,000 simulated values of the convergent sample, as appropriate. Under the described method we end up with predictive distributions for the future rates under full parameter uncertainty from which we extract all the essential summarising information required. We particularly examine future rates corresponding to ages 65, 75 and 85 over a fifty years ahead horizon. The plots of this section show the mean projections and the associated 95% confidence bands for the examined data-sets. The projections are compared in two manners. First, the BA-models are compared in Figures 6.8 and 6.9 for the two populations. In this case the models have been plotted to show the logit mortality rates, $q(x,t)$, for $x = 65, 75, 85$. Second, in Figures 6.10-6.13 we compare each extended model within the family it belongs to for the two populations.

The differences between the two BA-models follow closely those of the original LC and CBD models. For example, the BA-CBD model develops wider confidence bands as the projection age increases. As a result the intervals at old ages of the data-sets show marked differences between the two models for both populations. The higher uncertainty in the long-term dynamics of the EW data, noted in the previous section, also contributes to the increased spread of the predictive mortality rates distribution compared to Canada. For example, the confidence bands of the forecasted rates of the BA-CBD model for middle ages overlap those of the BA-LC in the case of the EW data, but not in the case of Canada. Another observation is that for the ages examined here the mean mortality trajectory of the BA-CBD model is usually greater than that of the BA-LC model. Finally, the projections of the BA-LC model for middle and high ages of the EW data-set develop noticeable lumpy trends due to distortions generated from the residuals process. These distortions are exactly due to the fact that the autoregressive model reproduces the mortality characteristics of the respective population. Once the cohorts of the initial population exit the age window of the projected mortality table, the forecasts become smoother due to the lack of that information. This will be even more evident in the examination of the mortality improvements in Section 6.5. Similar lumpiness in the mean forecasts of the BA-LC

model may also be observed in the case of the Canada data, although the feature is much less prominent therein.

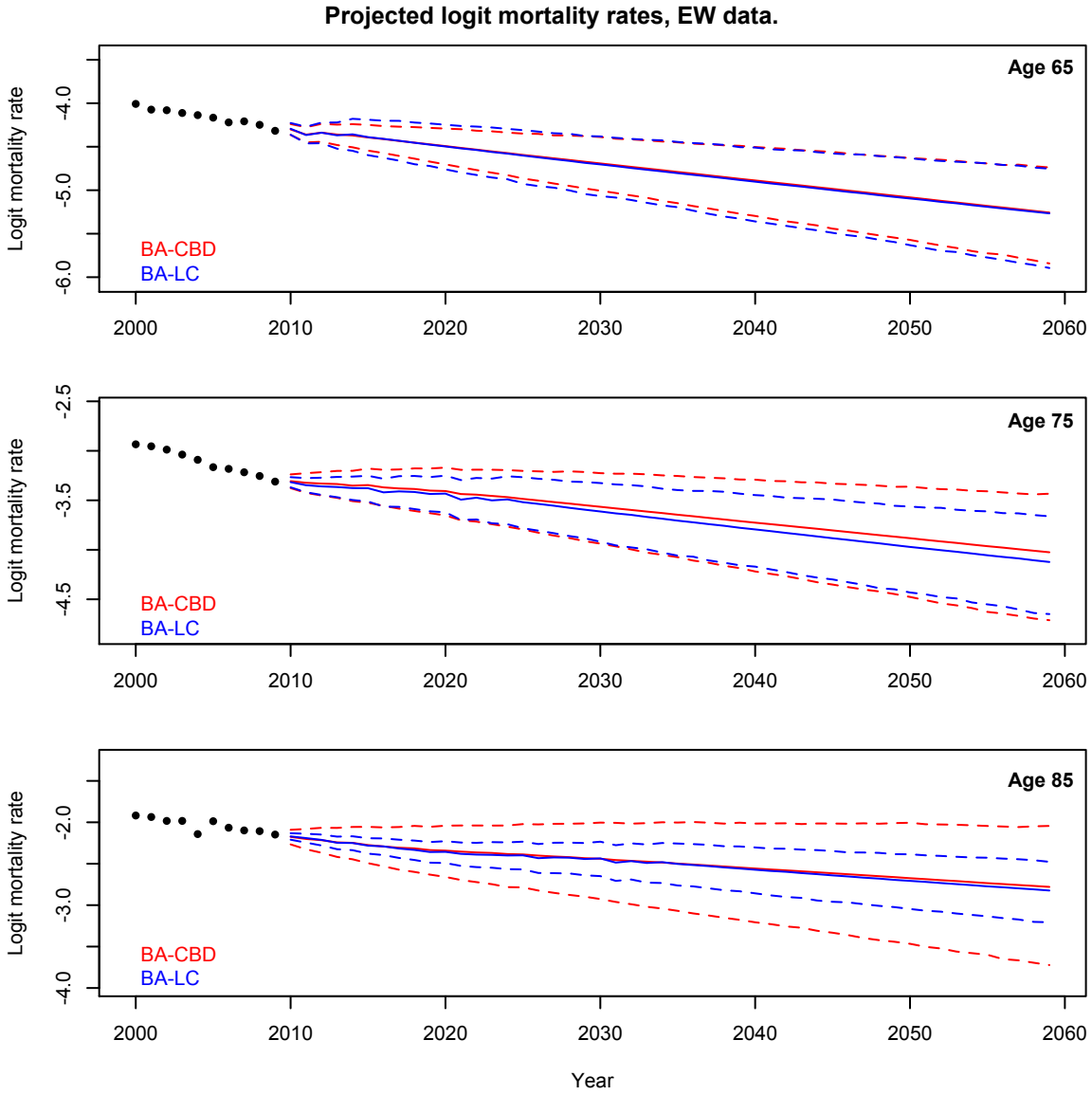

Figure 6.8: Forecasted mean mortality rates with their associated 95% confidence bands on the logit-scale for ages 65, 75 and 85 over a 50 years ahead horizon for the BA-CBD and BA-LC models, EW data.

Next, we compare the projections of the BA-models to those of the other members of each family they belong to. The results are examined for the two populations separately. Figures 6.10 and 6.11 show the results for the EW data-set, and Figures 6.12 and 6.13 illustrate the projections for the Canada data. The extreme improvements of the CBD and M7 for the first ages of the EW data observed in Section 6.1 are the reason behind the different levels of the forecasted mortality rates for age 65 in Figure 6.10. The BA-CBD serves as a middle solution between the other two models

of the family. For middle age projections of the EW data the thee models begin from very close to one another. However, beyond the first 15 projection years the mean trajectories clearly deviate. Model M7 yields the higher mean rates, followed by the BA-CBD and CBD models respectively. The confidence bands of the BA-CBD overlap those of the other two models after around the same time instance. Finally, the mean forecasts of the three models are very close for around the first 20 years of the projection period when examining the bottom panel of Figure 6.10.

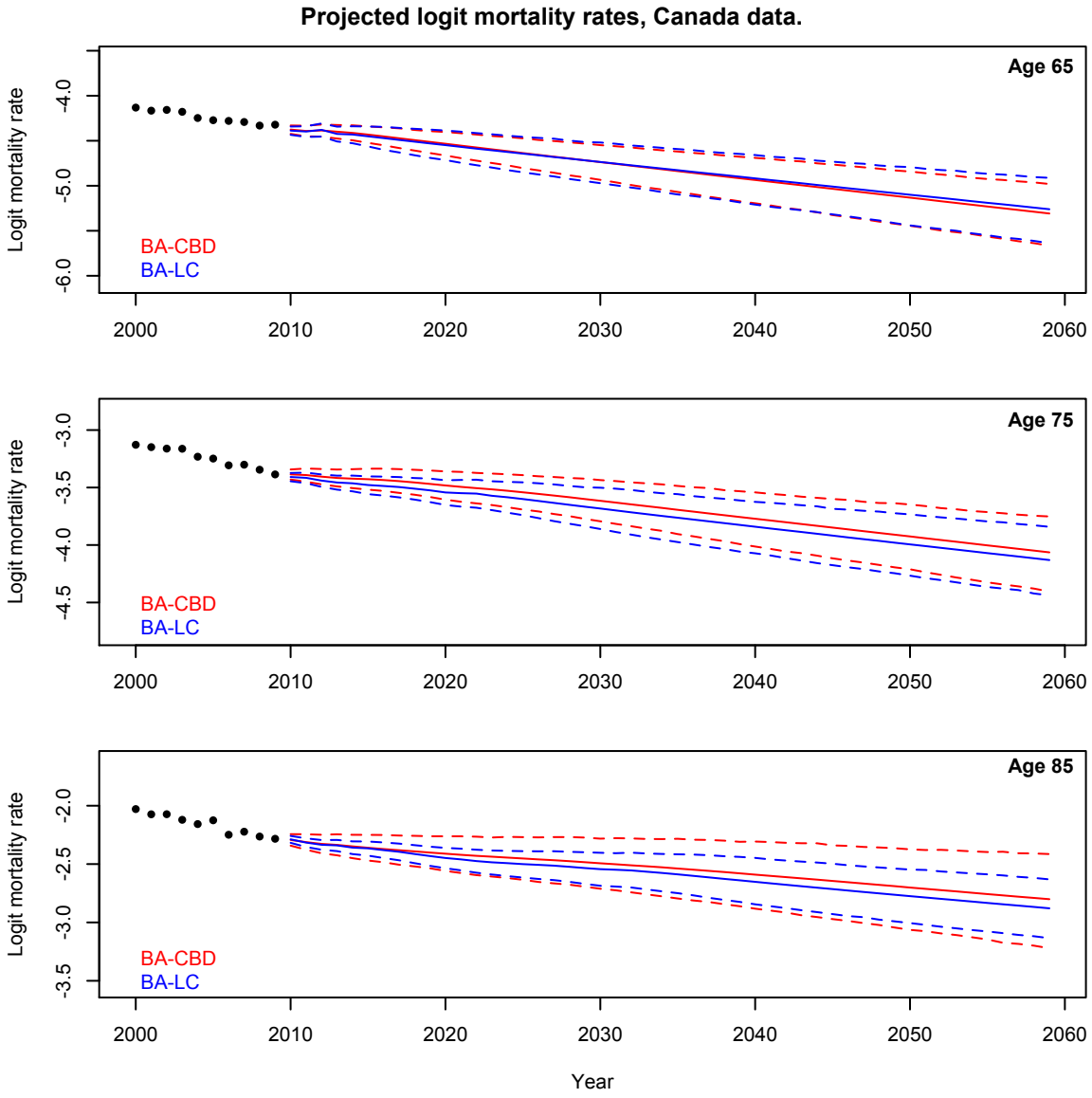

Figure 6.9: Forecasted mean mortality rates with their associated 95% confidence bands on the logit-scale for ages 65, 75 and 85 over a 50 years ahead horizon for the BA-CBD and BA-LC models, Canada data.

As soon as smooth, projected cohort effects start contributing to the construction of the rates under model M7, its mean forecasts lump upwards, while those of the CBD and BA-CBD models stay aligned. The different trend produced by the forecasts of model M7 also leads to remarkably different lower percentiles for the future distributions of the mortality rates. Lastly, as the examined age increases, the width of the confidence bands of the BA-CBD gets much greater than those of the original CBD model, signifying the contribution of the additional residuals model in the uncertainty levels about the future mortality rates. Since the full Bayesian solution of the original CBD model only slightly alters the confidence level in the relevant projections as shown in Section 2.4.1, the impact of the residuals augmentation is evident.

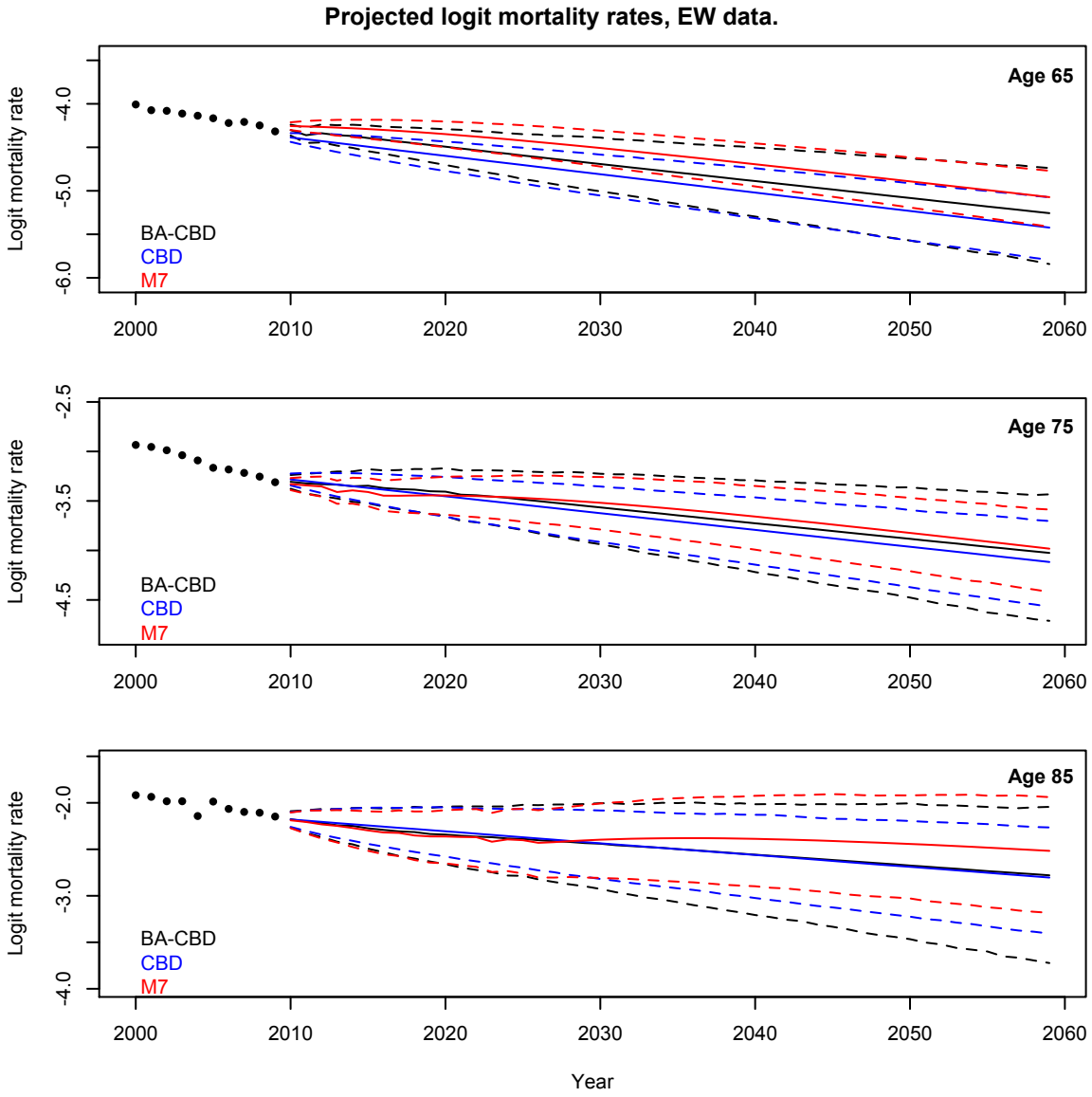

Figure 6.10: Forecasted mean mortality rates with their associated 95% confidence bands on the logit-scale for ages 65, 75 and 85 over a 50 years ahead horizon for the CBD, M7 and BA-CBD models, EW data.

Moving to the LC family, the projections of the APC model start from significantly lower points compared to the other two models at almost all examined ages. Therefore, we observe that for middle and old ages of the EW data-set the projections of the LC and the BA-LC models are very close. The residuals augmentation of course yields again greater confidence bands for the BA-LC model. The first few projected rates for age 65 are very close between the LC and APC models. Surprisingly, in the long term the forecasts of APC approach those of the BA-LC while the LC outcome is found significantly below them. Essentially, the models that take into account the cohort effect agree.

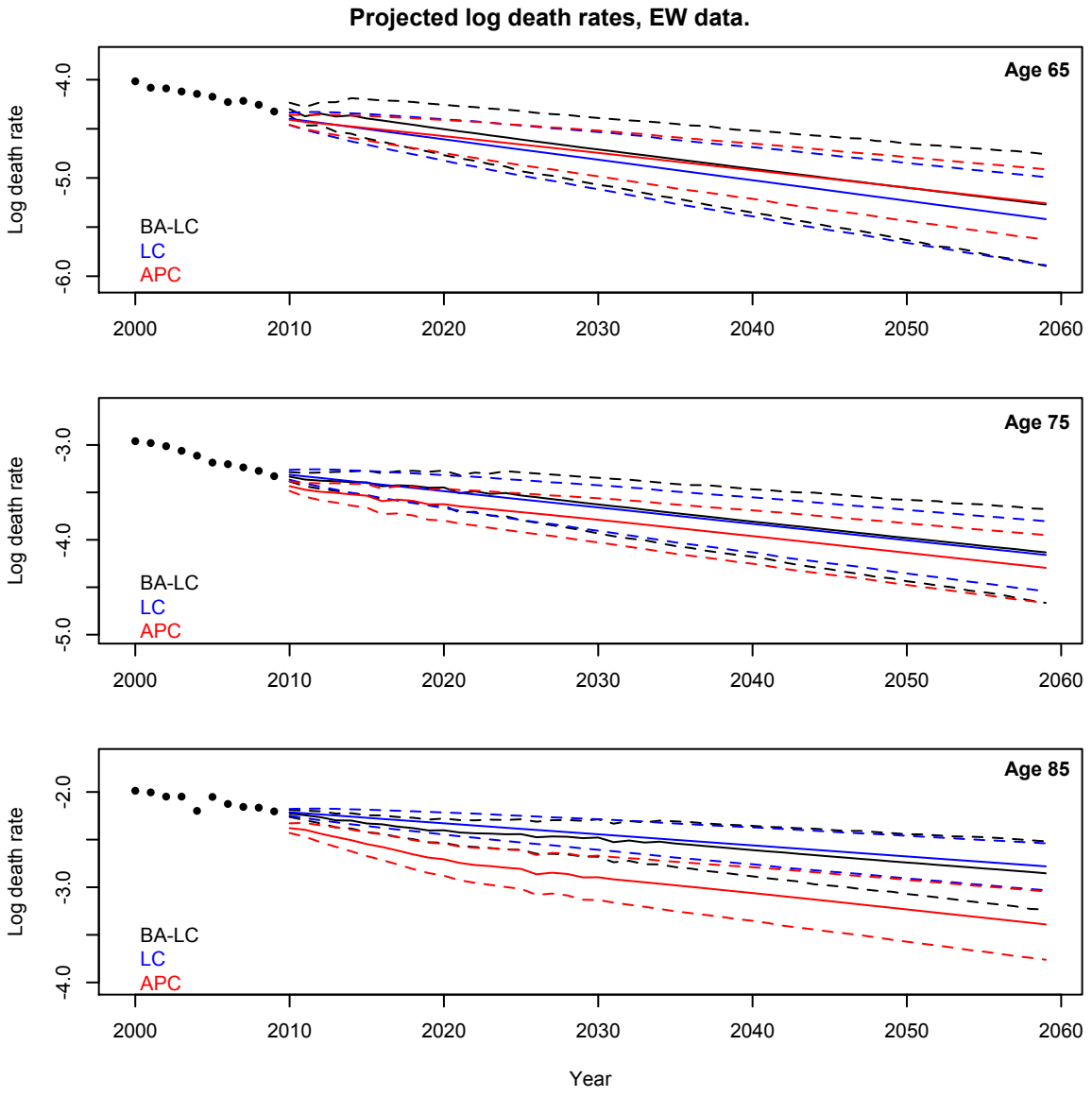

Figure 6.11: Forecasted mean death rates with their associated 95% confidence bands on the log-scale for ages 65, 75 and 85 over a 50 years ahead horizon for the LC, APC and BA-LC models, EW data.

The projections of the Canadian data-set mainly reveal the lower level of uncertainty inherited therein for both families of models. The forecasts of the CBD family for young ages of the data-set produce mean future rates that are distinguishably higher for model M7. The opposite is true for middle ages, where model M7 predicts evidently lower mean trajectories compared to the other two members of the family. Lastly, for old ages the three models seem to agree except for a small period during which model M7 starts incorporating projected cohort factors.

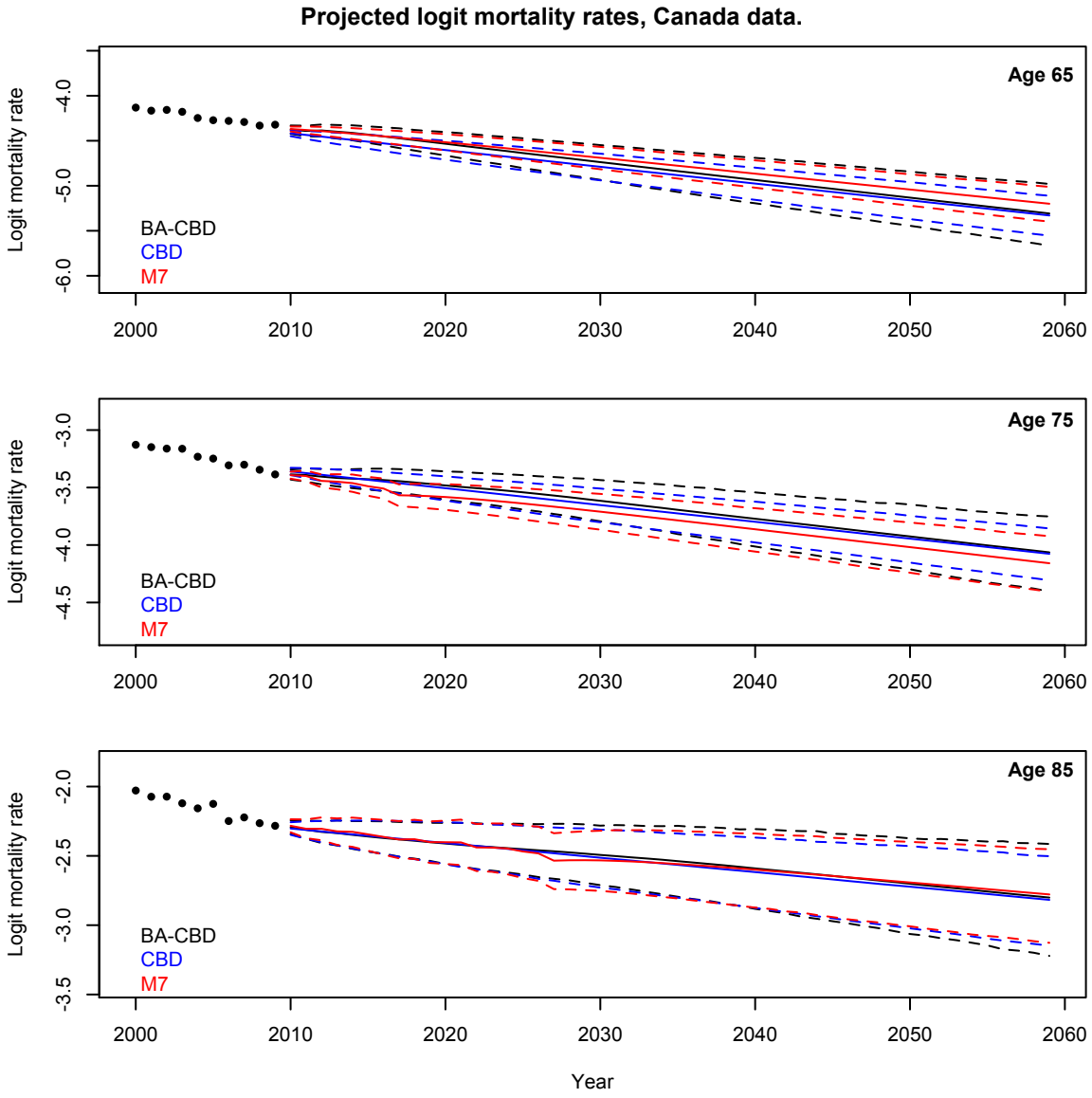

Figure 6.12: Forecasted mean mortality rates with their associated 95% confidence bands on the logit-scale for ages 65, 75 and 85 over a 50 years ahead horizon for the CBD, M7 and BA-CBD models, Canada data.

The difference in the levels of uncertainty in the Canadian population is more evident for middle and old ages of the data-set. While in the case of the EW data the confidence bands of the BA-CBD model overwhelmingly overlap those of the other two models, that only happens for the very long term projections of old ages in the Canada data. The mean forecasted rates of the LC family of models on the other hand differ significantly relative to one another for the various ages examined. The mean projections of the BA-LC model almost coincide with those of the LC model for middle ages of the data-set. As was the case with the EW data-set, the APC model develops significantly different trends for the central death rates of older ages of the data.

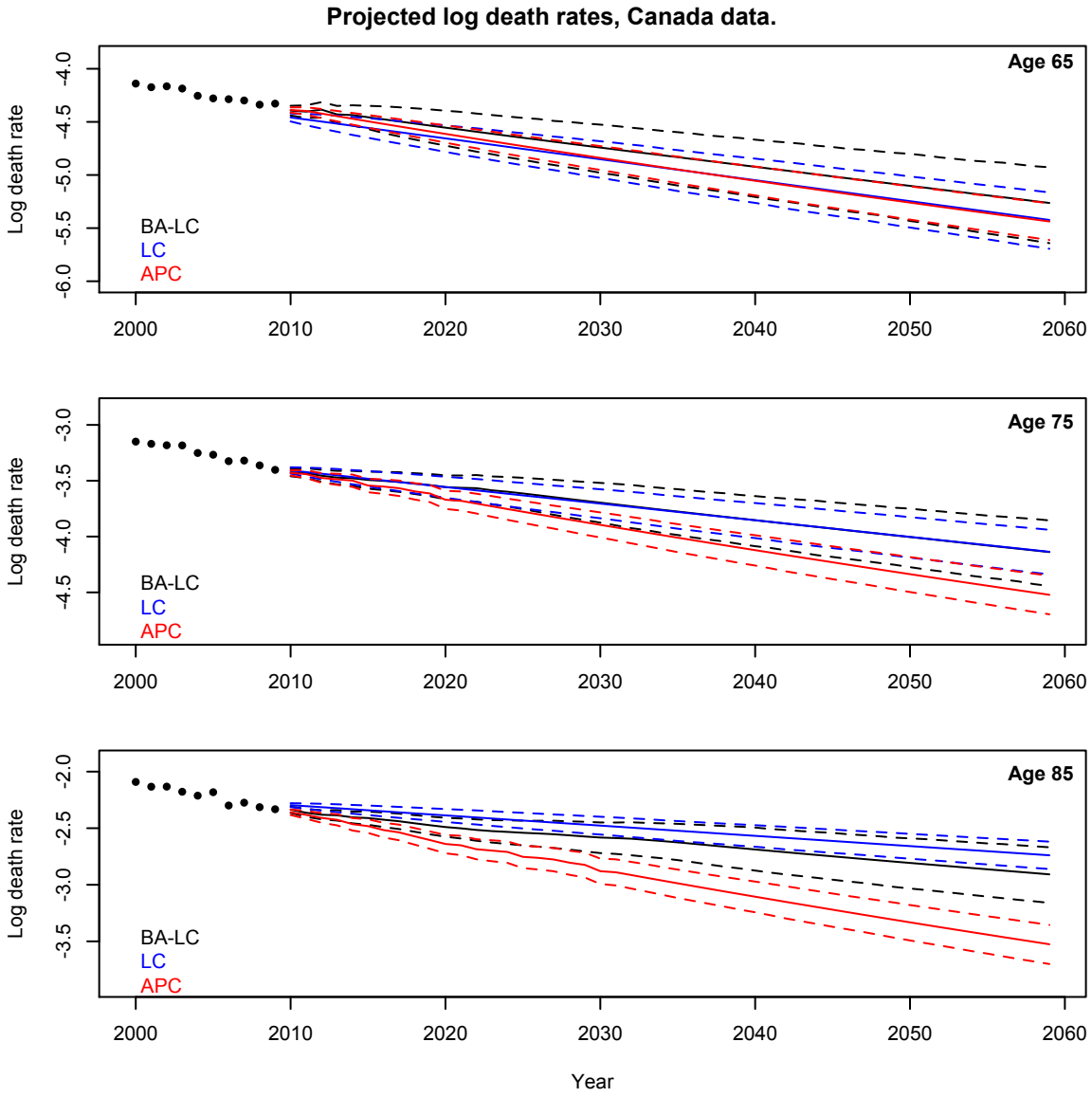

Figure 6.13: Forecasted mean death rates with their associated 95% confidence bands on the log-scale for ages 65, 75 and 85 over a 50 years ahead horizon for the LC, APC and BA-LC models, Canada data.

6.5 Mortality metrics

The projections carried out in Section 6.4 are used here to calculate relevant mortality indices in order to examine the behaviour of the BA-models. The results are also compared to those of the mortality models of Chapter 1. Additionally, the forecasts of all mortality models are judged according to their performance relative to the realised mortality rate, where this is available. Finally, the improvement rates implied by the forecasts of the mortality models are examined.

Based on the projected mortality tables the posterior distribution of the survivor index $S(65, t)$ over the next 25 years is extracted. The results are plotted in Figure 6.14 according to the family each extension belongs to, and along with the rest members of that family for both data-sets. The BA-CBD model projection reveals greater downside risk for the survival probability of the examined cohort in both datasets. By the end of the projection period the confidence bands of the BA-CBD model cover those of the other two models in the case of the EW data. The mean forecast of the BA-CBD model for the Canada population is much closer to that of the original CBD model for the majority of the projection horizon. However, by year 2034 the three models appear to converge.

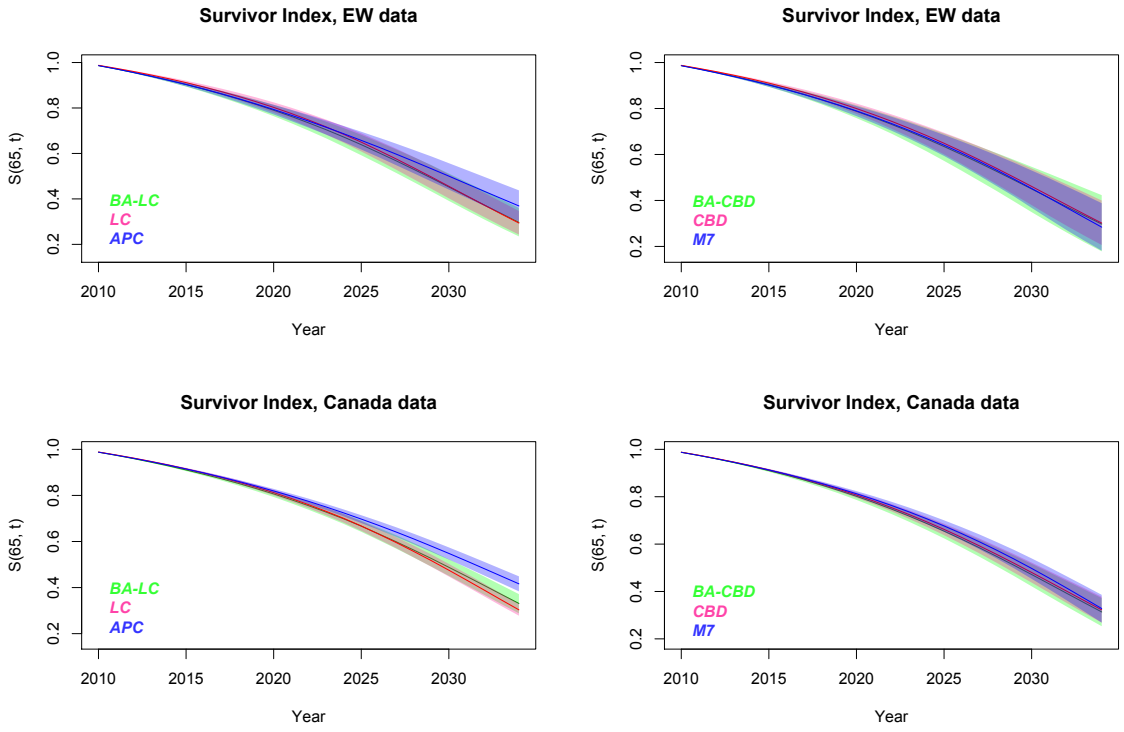

Figure 6.14: Projections of survivor indices, S(65, t), and associated 95% confidence bands over the next 25 years for the BA-CBD, BA-LC, LC, APC, CBD and M7 mortality models, for the EW and Canada data-sets.

The same increased level of downside risk is observed in the case of the EW data for the BA-LC extension. Also, the confidence bands of the BA-LC model cover those of the LC model by the end of the projections period in that case. The forecast of the BA-LC model falls between those of the LC and the APC models for the Canadian population, while the produced confidence interval is very narrow too.

Based on the projection of the survivor index, the predictive distribution of a unit term annuity value payable in arrears for the following 25 years starting in 2010 under a constant 4% interest rate is calculated. Similarly to Section 1.4.3, the relevant cohort life-expectancies are also considered. The densities of the respective distributions, along with those of the remaining mortality models, for both populations are shown in Figure 6.15. Beyond the notable difference of the APC model, the other models produce very similar results. Probably the most notable impact is for the life expectancies of the Canada data, where the distribution is noticeably shifted. The BA-models generally yield fatter tails and distributions with greater range than the non-augmented models. Characteristically, the life expectancy density of the BA-CBD model for the EW data overlaps all other densities in the top right panel of Figure 6.15.

As in Chapter 1, the period life-expectancies at age 65 are calculated from the projected mortality table for years 2034 and 2059. The results for the mean and standard deviation of the relevant distributions are included in Table 6.3 and are compared to Table 1.4 of Section 1.4.3. The mean forecasts are generally higher under the BA-model compared to the other members of their respective family. Hence, the extended models generally produce lower mortality rates. However, more notable are the increased uncertainty levels of the augmented models. In particular, the BA-CBD model yields standard deviations for the period life-expectancies which are 33% and 90% greater than those of model M7 for years 2034 and 2059 respectively for the EW data. The increase is also apparent for the Canada population but of a much lower magnitude.

In the time between the preparation and the completion of the current work the HMD has updated its data of the EW population for years 2010 and 2011. Given the predictions of the mortality models, we can examine the performance our forecasted rates in contrast to what has been realised. Figure 6.16 illustrates the projected densities of the central death rates, $m(x, 2010)$ for $x = 65, 75, 85$ along with the realised rate at that year. The projection of the BA-models yield forecasted rates very close to the realised rate for age 65, although both the original LC and CBD models perform poorly. For age 75 the BA-models are concentrated around the realised value as the majority of models do. Finally, the BA-LC model forecasts very accurately the central death rate for age 85, in line with the results of the LC model.

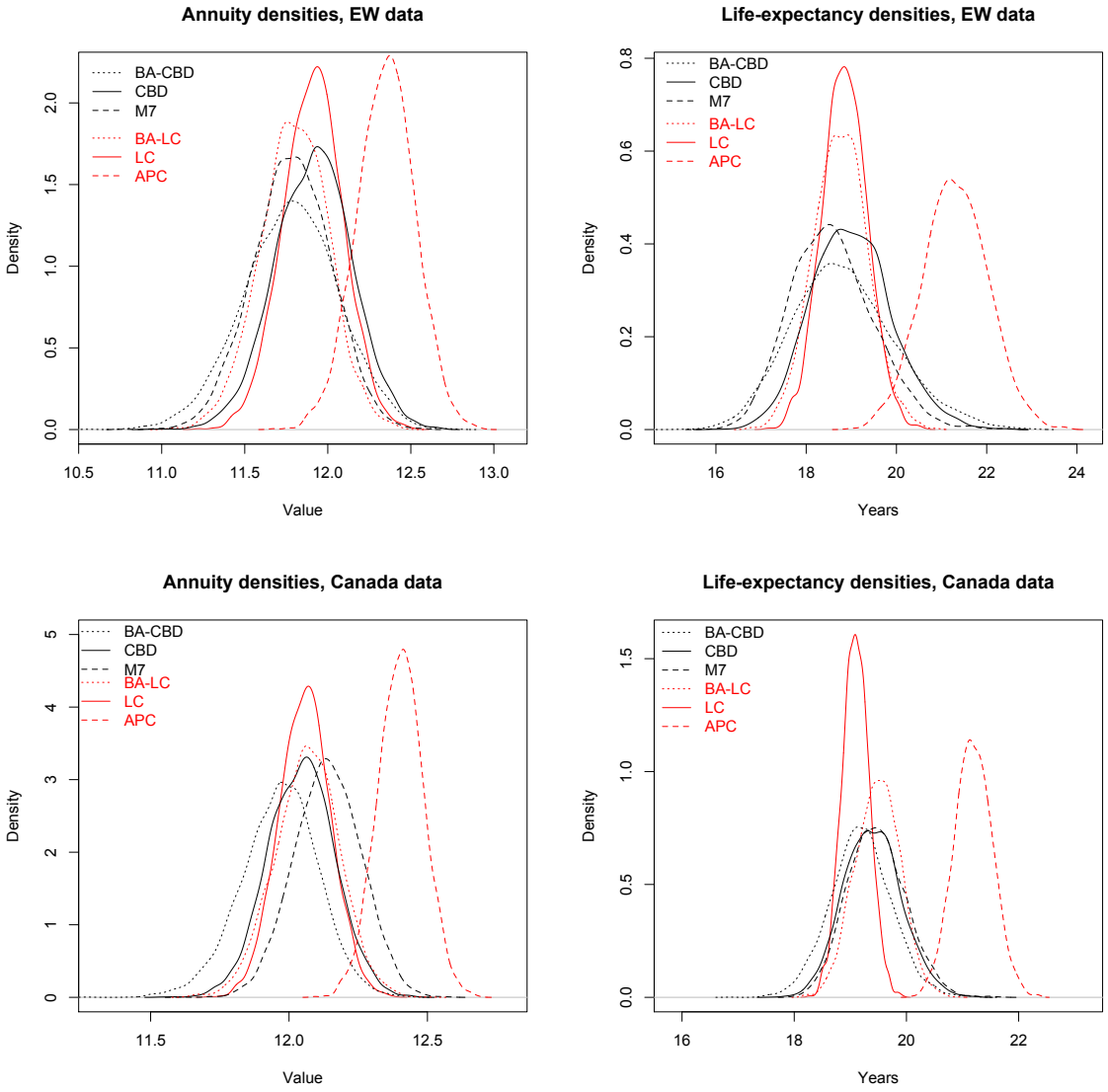

Figure 6.15: Density plots of unit 25 years term-annuities payable in arrears starting in 2010 under a constant 4% interest rate and cohort life-expectancies for individuals aged 65 in 2010 for the BA-CBD, BA-LC, LC, APC, CBD and M7 mortality models, for the EW and Canada data-sets.

	EW data				Canada data			
	Year 2034		Year 2059		Year 2034		Year 2059	
Model	$\mathbb{E}(L)$	$Sd(L)$	$\mathbb{E}(L)$	$Sd(L)$	$\mathbb{E}(L)$	$Sd(L)$	$\mathbb{E}(L)$	$Sd(L)$
BA-CBD	22.52	2.5054	27.45	5.9597	22.57	1.1453	26.16	2.3693
BA-LC	22.36	1.2313	26.50	2.3367	22.99	0.8353	26.72	1.5749

Table 6.3: Summary of period life-expectancy distributions at age 65 in calendar years 2034 and 2059 for the BA-CBD and BA-LC models, EW and Canada data.

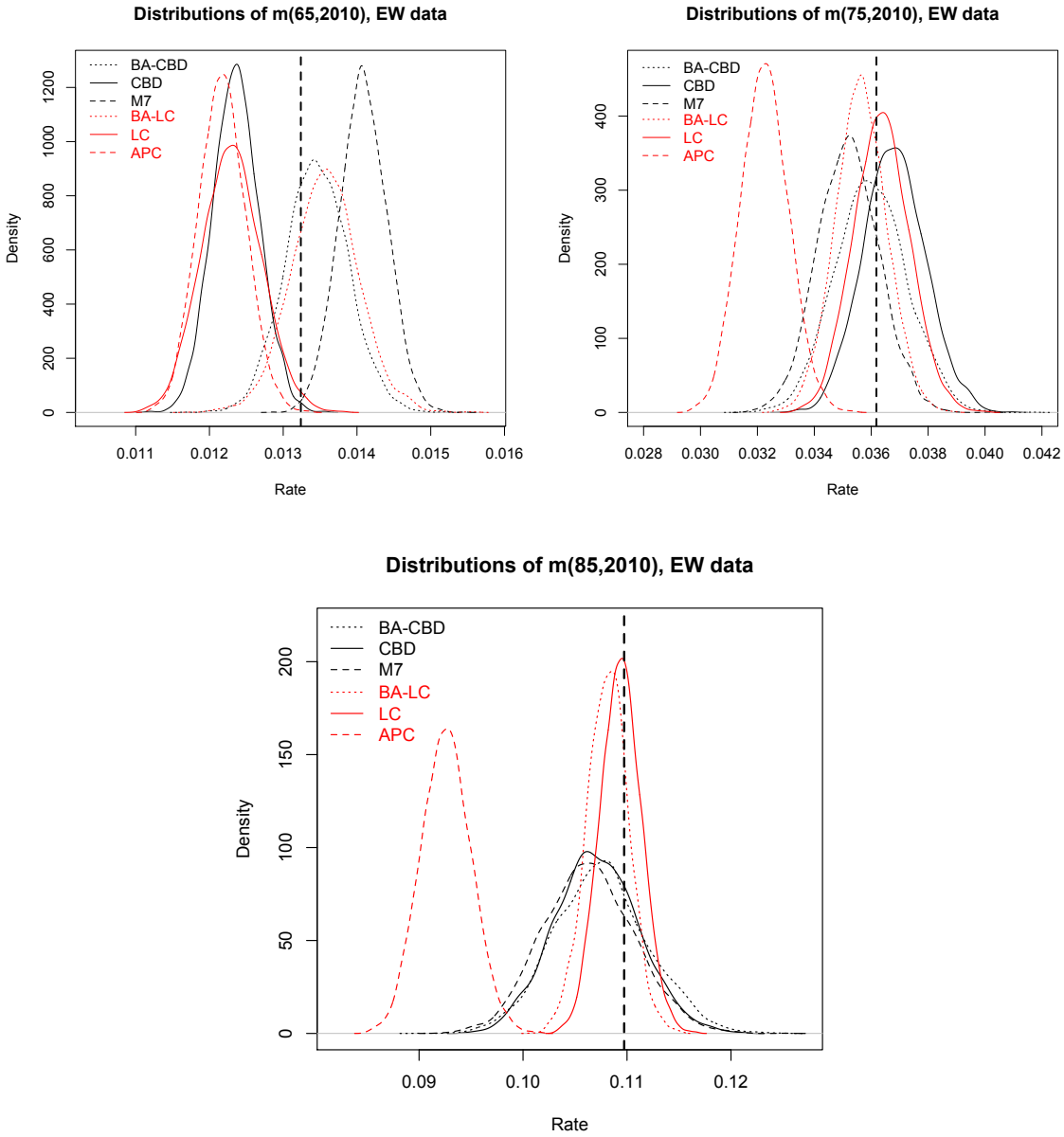

Figure 6.16: Forecasted densities of the central death rates, $m(x, 2010)$ for $x = 65, 75, 85$ under the BA-CBD, BA-LC, LC, APC, CBD and M7 mortality models, EW data. The dashed vertical line in each plot indicates the realised rate during 2010.

In summary, the BA-models seem to lead the forecasts towards the realised death rates in cases where the other members of each family perform inaccurately. At the same time, the augmentation does not significantly alter the predictions when the projections of the mortality models fall close to the observed rate.

Finally, we examine the forecasted mortality improvement rates produced by all the stochastic models. First, the distributions of improvements for the fixed ages 65, 75 and 85 are assessed. Second, we look at the matrices of mean improvement rates for all ages across the whole projection period. If the projection is conducted for time

instances t_j, the improvement rate of the mortality rate, $q(x,t)$ for fixed age x, is calculated as:

$$\frac{q(x,t_j-1)-q(x,t_j)}{q(x,t_j-1)}.$$

The results, adjusted to percentage forms, for all implemented mortality models are plotted in Figures 6.17 and 6.18 for the EW and Canada data, respectively. The graphs show the mean forecasted improvement rate and the associated 95% confidence bands. In each plot, an additional line, indicating the zero improvement rate, has been added as a benchmark. The right panels refer to the Canadian population and the left panels show the results for the EW data. The figures are plotted in the same scale for the two populations to emphasise once more the difference in the uncertainty underlying the two data-sets. Since we have forecasted the mortality rates up until year 2059, the above formula for the improvement rates implies we have the improvement rates for years 2009, 2010, ..., 2058.

Based on the plots of Figures 6.17 and 6.18 several comments may be made. First, all models predict positive improvements for the mortality rate over the following 50 years. Also, the rate of improvement appears to converge at some constant level after the first years of the forecasting period. The convergence is achieved earlier for younger ages commonly for both data-sets. The improvement rates are generally very close between the BA-CBD, BA-LC, CBD and LC models for the EW population, amounting to around 1.2% per year. In contrast, the cohort-extended models, APC and M7, produce distinctive results. On the one hand the long term improvement rate of the APC model is around 1.7% and on the other hand, that of model M7 is around 0.65%. The deviations are more extreme when considering the Canadian population. The BA-CBD and CBD models yield improvement rates of around 1%, with the BA-LC returning rates of 1.2% and the LC and M7 models giving figures close to 0.85%. Finally, the APC model again gives the most optimistic predictions with average annual improvements rates of 2.2%. More interesting than the mean improvement rates are the associated confidence levels of these improvements. Interestingly, the BA-LC is the model which returns the widest confidence bands for the future improvement rates when considering the EW population. The observed differences are extreme for young ages and lessen as the examined age increases. However, given the single long-term stochastic factor of the BA-LC, the results remain questionable. The uncertainty in the forecasted improvements is heavily reduced in the case of the Canada data. This is stressed even more by the identical scale of the graphs for the two populations. For young ages, the BA-LC still produces the most uncertain mortality improvements, but the differences with the rest of the models are evidently reduced. The pattern changes as the examined age increases, so that the BA-CBD model is the one that returns the widest confidence bands for the predicted improvements.

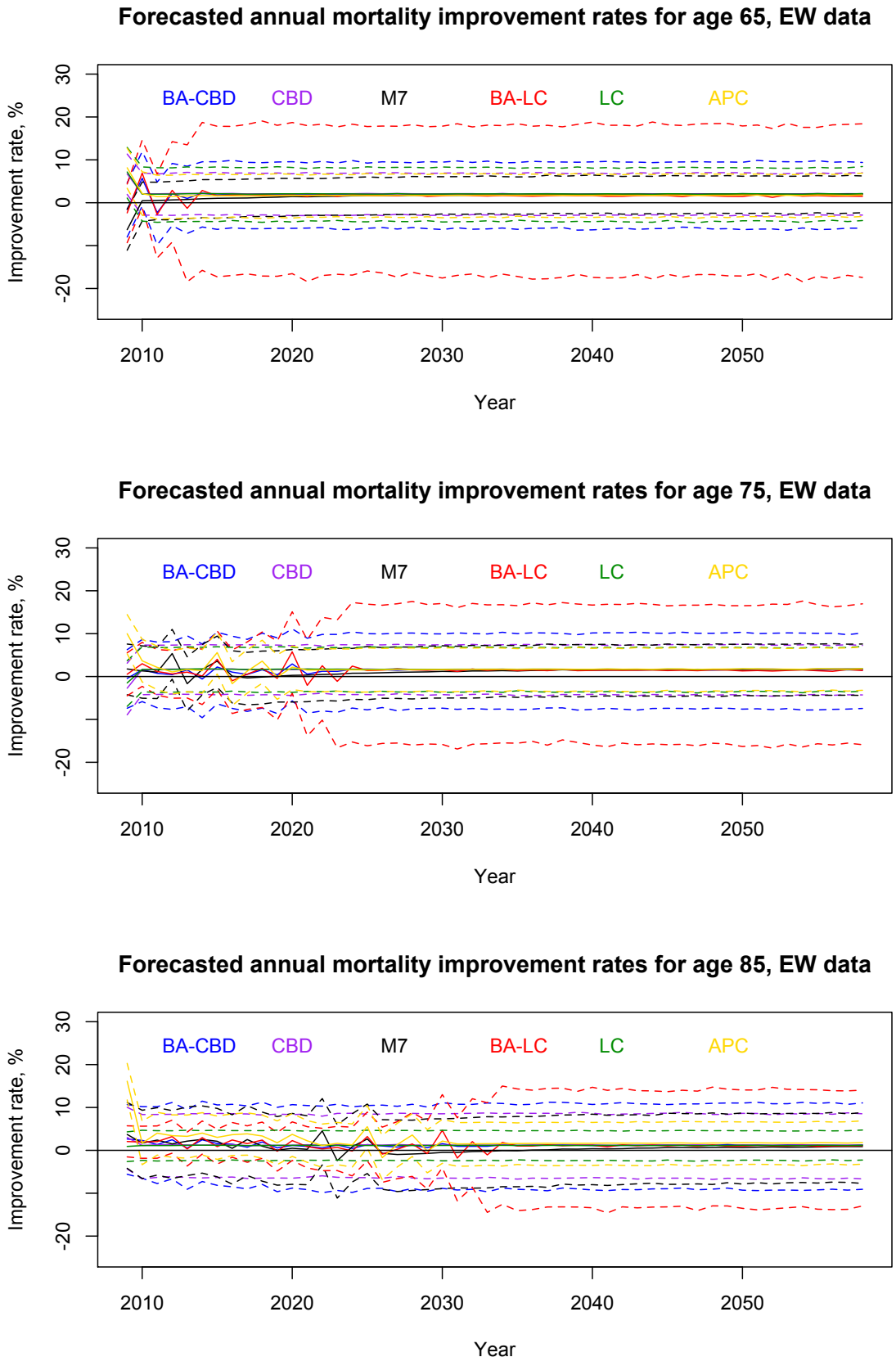

Figure 6.17: Projected mean improvement rates of $q(x,t)$ for $x = 65, 75, 85$ and associated 95% confidence bands over the period from 2009 to 2058, EW data.

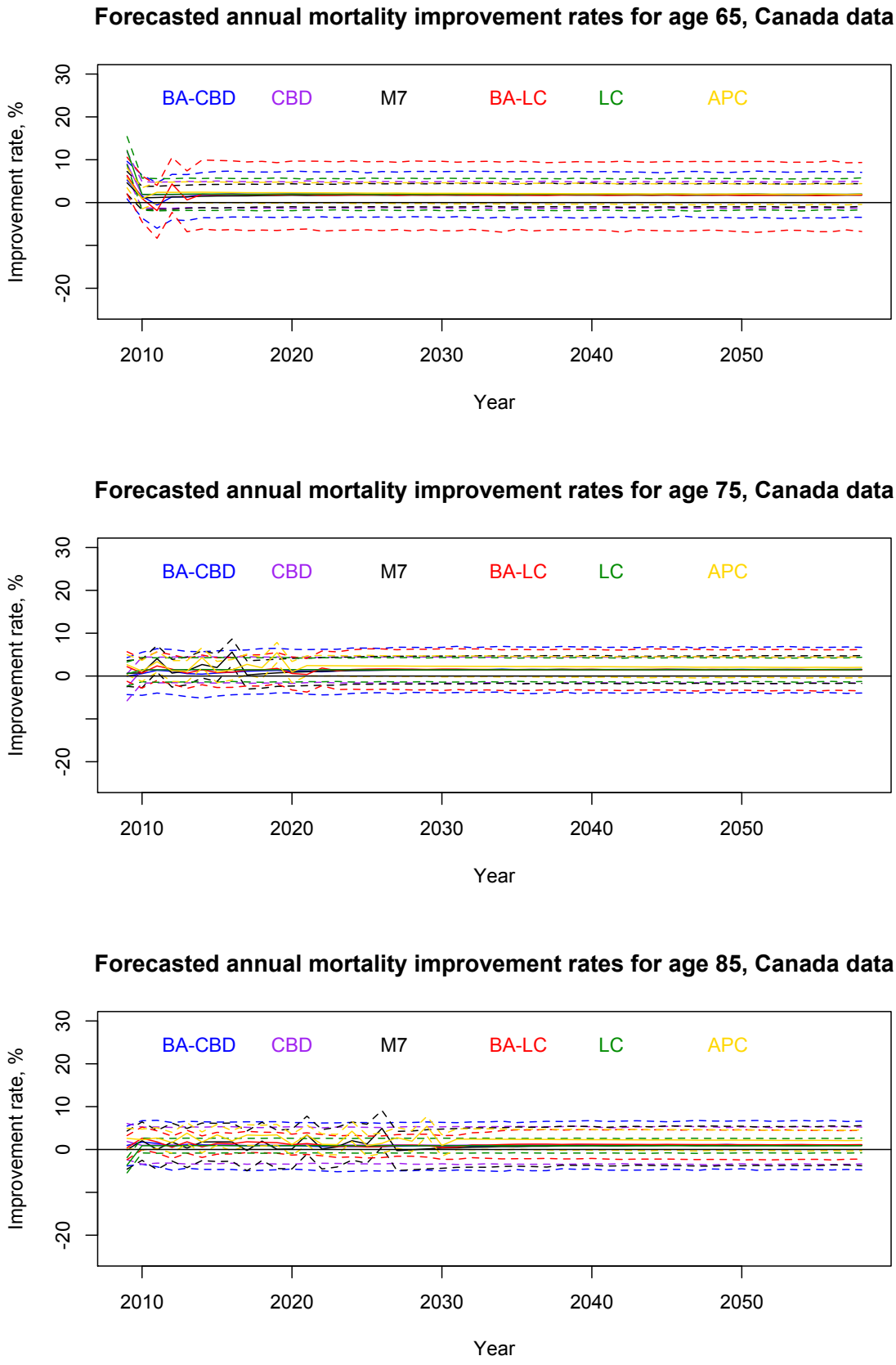

Figure 6.18: Projected mean improvement rates of $q(x,t)$ for $x = 65, 75, 85$ and associated 95% confidence bands over the period from 2009 to 2058, Canada data.

The matrices of the mean projected improvements across all ages for all stochastic mortality models and for both populations are shown in Figures 6.19 - 6.22. The left panels of each plot show the improvement rates for the full projection period except for the first, as commented in Section 6.1. The right panels display only the final 20 years of the forecast. The projections of the LC and CBD models are identical for the two cases just described. However, the models which incorporate some sort of cohort structure produce distinctive results. It appears that improvement rates are more extreme for cohorts included in the fitting data-set. For new entrants in the projection, the mean forecasted mortality improvements have commonly smaller range and display less persistent patterns. For example, model M7 seems the one which produces the less patterned plots for both populations. This is most probably an effect of the information carried by the data and fed into the forecasts through the residuals model. Commonly, the long term forecasts generally return decreasing mean improvements with age. This is the case for all graphs except for that of the APC model for the Canada population, where the improvements are increasing in line with age. The projections of the APC model are also suspicious in the case of the EW data where we observe very similar mean improvement rates for all ages, with years of commonly high improvements followed by years of commonly low rates and vice versa. The BA-models project long term improvements that are very similar to those of the fitted cohorts of the data-sets. This is particularly the case for the BA-LC model, where the diagonal structures are reproduced under the correct order, but in lower magnitude compared to the first 30 years. The BA-CBD seems to create cohort patterns mainly in the EW data projection. In the case of the Canada population, the long term forecasts lose that diagonal shape and develop waving patterns.

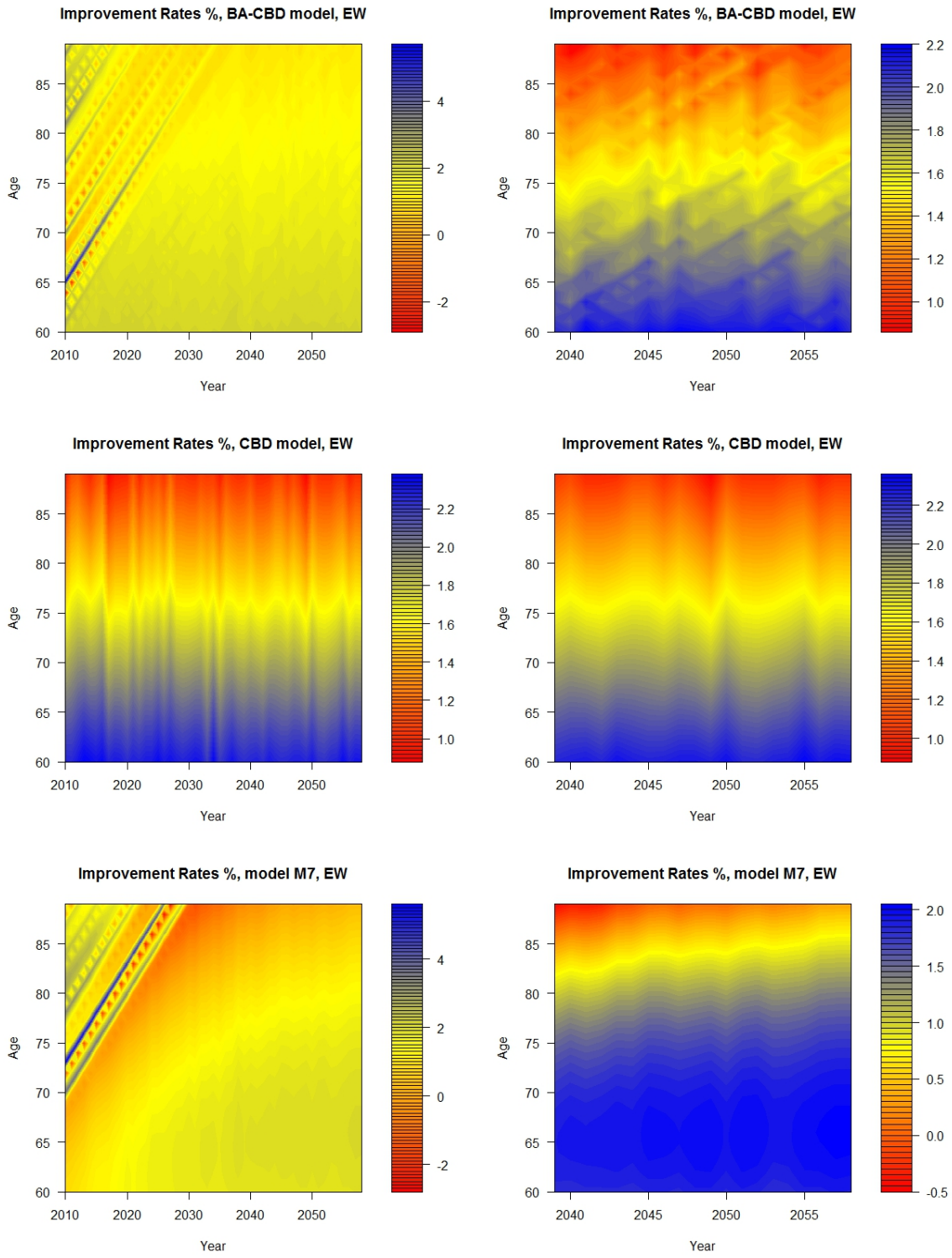

Figure 6.19: Projected mean improvement rates of $q(x,t)$ for all ages, $x = 60, \ldots, 89$ for the BA-CBD, CBD and M7 models, EW data. The left and right panels show the improvements across the full projection period and during the final 20 years, respectively.

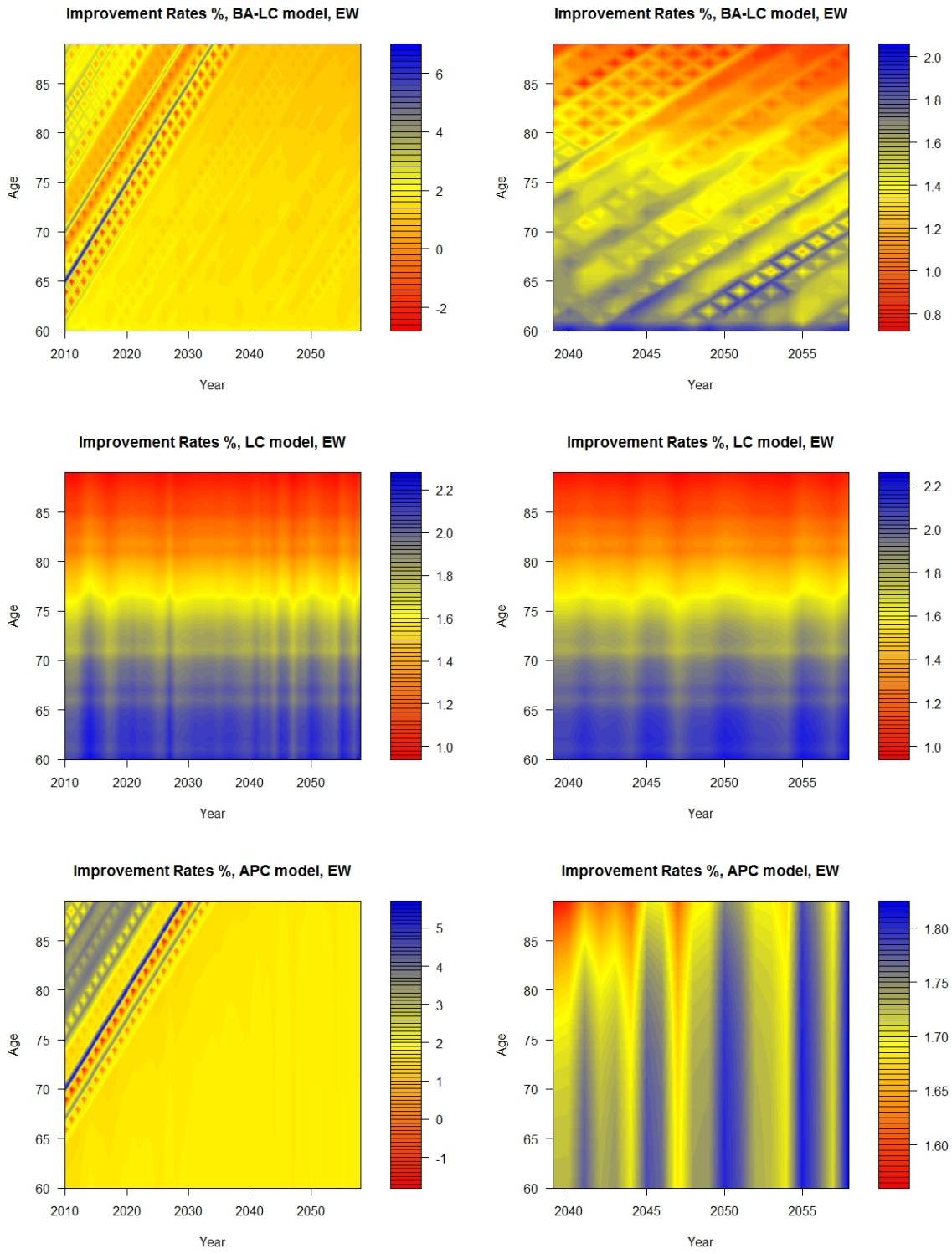

Figure 6.20: Projected mean improvement rates of $q(x,t)$ for all ages, $x = 60, \ldots, 89$ for the BA-LC, LC and APC models, EW data. The left and right panels show the improvements across the full projection period and during the final 20 years, respectively.

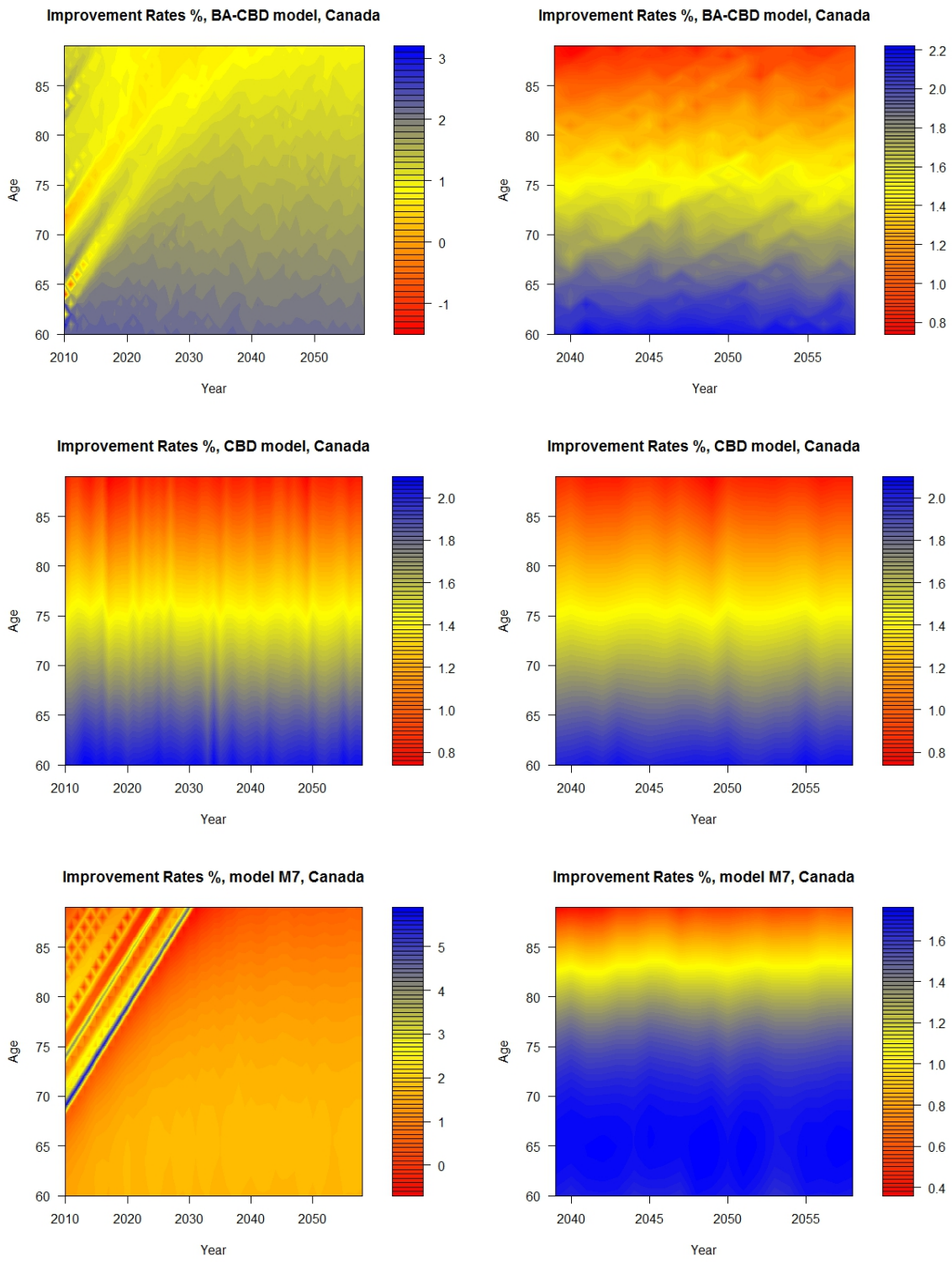

Figure 6.21: Projected mean improvement rates of $q(x,t)$ for all ages, $x = 60, \ldots, 89$ for the BA-CBD, CBD and M7 models, Canada data. The left and right panels show the improvements across the full projection period and during the final 20 years, respectively.

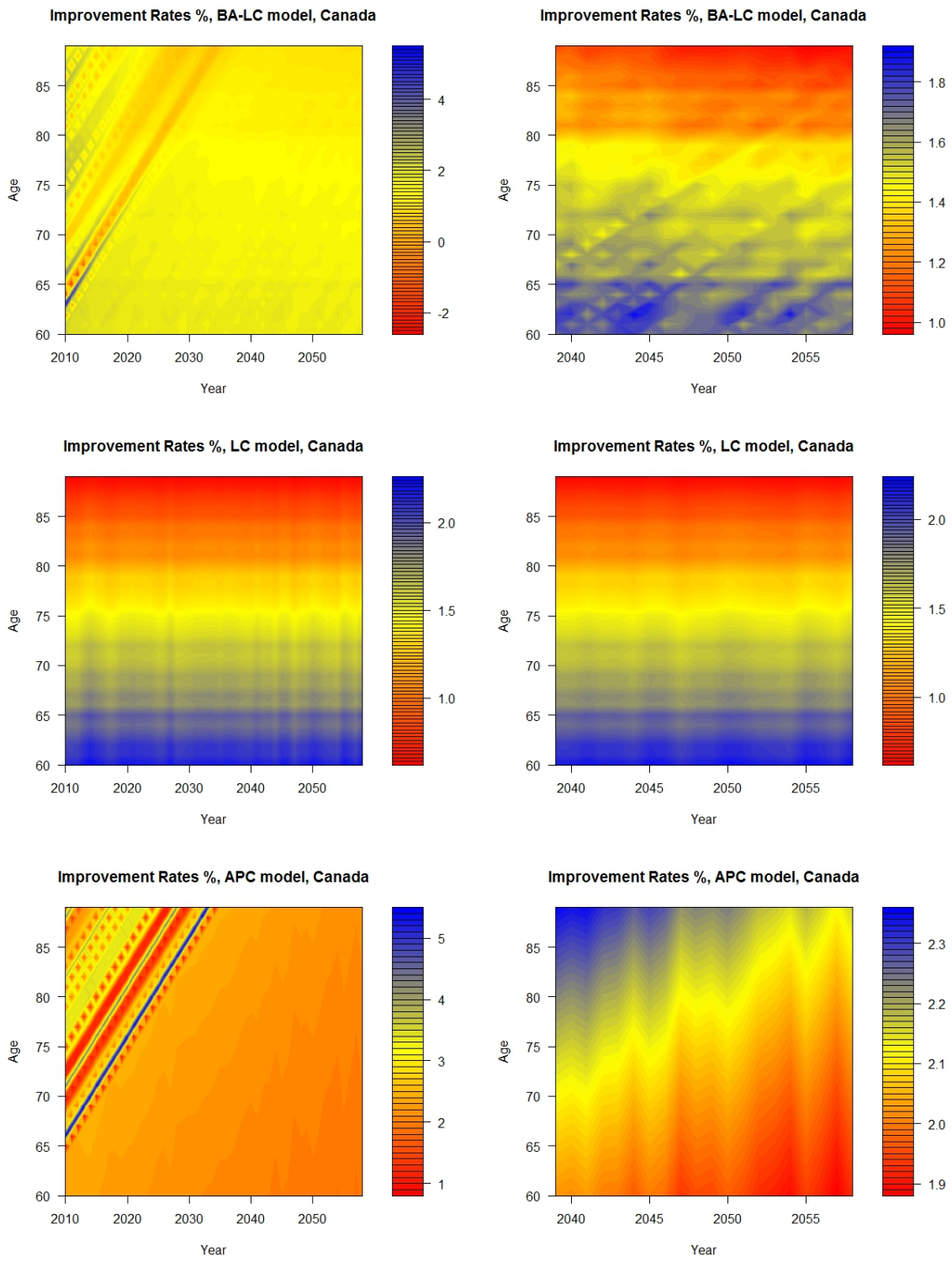

Figure 6.22: Projected mean improvement rates of $q(x,t)$ for all ages, $x = 60, \ldots, 89$ for the BA-LC, LC and APC models, Canada data. The left and right panels show the improvements across the full projection period and during the final 20 years, respectively.

6.6 Robustness

This Section is aimed at examining the robustness properties of the BA-models. In order to assess the consistent interpretation of the parameters, the models are fitted to a subset of the original data-set. In particular, the EW population for ages 60-89 and for years 1960-1999 is used. Hence, the models are estimated for ten years less compared to Chapters 4 and 5, allowing for examining the restricted set of period effects, the behaviour of the residuals model and for in-sample performance of the forecasts.

First, the BA-CBD model is examined. The top two panels of Figure 6.23 show the mean posterior estimates and the associated 95% credible bands for the series of period effects under the full and restricted estimation in black and red lines, respectively. The main characteristics of the posterior distributions are maintained, so that $\kappa_t^{(1)}$ is subject to much less uncertainty compared to $\kappa_t^{(2)}$. The two sets are slightly shifted under the restricted estimation scheme, but the trend of the processes seems little affected. Thus, when examining the posterior distributions of the parameters of the random walk model for the period effects, $\boldsymbol{\delta}$ and V_ζ, they also appear very close to the results of the full estimation. More specifically, parameter σ_1^2 yields almost the same posterior form, since the variability of $\kappa_t^{(1)}$ during the last decade is minimal. In contrast, the posterior distributions of σ_2^2 and σ_{12} are more affected, developing heavier right tails under the restricted estimation. Accordingly, the distribution of the drift δ_1 is shifted towards higher values, since the steep evolution of the last ten $\kappa_t^{(1)}$ values are omitted. Finally, the posterior distribution of δ_2 remains very similar to that produced by the estimation under the full data-set.

The middle and bottom panels of Figure 6.23 display the main components of the residuals model. The graphs show the posterior densities of the autoregression parameters, α_4, α_5 and α_6, which are involved in the construction of the residuals states for the majority of the ages of the data, and of the variance parameter of the covariance matrix V_Z. Parameter α_5 is the one that is most affected by using fewer observation years. The resulting posterior distribution is centred around lower values implying weaker cohort impact when less years are used to estimate the BA-CBD model. Parameters α_4 and α_6 are only slightly shifted compared to the estimation under the full data-set. In all cases, the autoregression parameters develop posterior distributions with somewhat greater variance. The same is true for the variance parameter, v, on the diagonal of the covariance matrix, V_Z, which is also centred around 20% greater values under the calibration of this section. Overall, the increased posterior variance of the parameters of the residuals model is probably due to the limited information provided by the restricted data-set. However, the main characteristics of the residuals model are kept consistent.

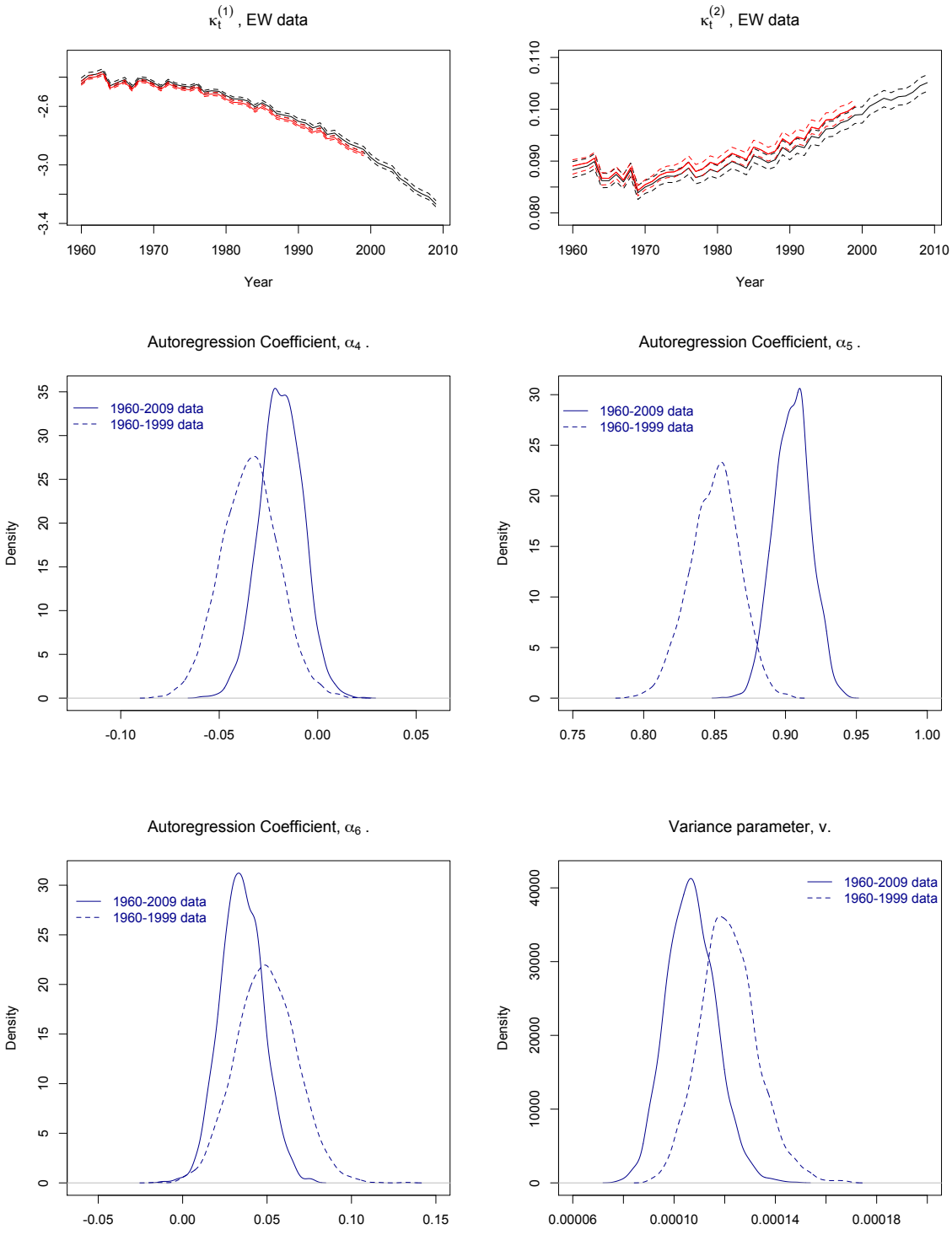

Figure 6.23: Top panel: Posterior distributions of the period effects series, $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$, of the BA-CBD model for the EW data, using data 1960-2009 and 1960-1999, in black and red lines respectively. Solid lines indicate mean estimates and dashed lines show the 95% credible intervals.
Middle and bottom panels: Posterior densities of the autoregression coefficients, α_4, α_5 and α_6, and of the variance parameter, v, of the BA-CBD model using the full and truncated EW data.

Next, we examine the behaviour of the BA-LC model after estimating the model with the restricted EW data-set. Figure 6.24 shows the original and the revised posterior distributions of the age effects series, $\beta_x^{(1)}$ and $\beta_x^{(2)}$, and of the period effects series, $\kappa_t^{(2)}$, of the BA-LC model. Parameters $\beta_x^{(1)}$ are mainly affected for the first and last few years of the data-set. However, their deviation is minimal and their credibility bands are narrow, similarly to the estimation under the full EW data. Analogous conclusions might be drawn for the period effects series, $\kappa_t^{(2)}$. The relevant posterior distributions are only slightly shifted upwards after the first 10 years of the data-set, while the magnitude of the resulting credibility intervals is roughly kept at the same levels. The form of the posterior distributions of $\kappa_t^{(2)}$ also determine the characteristics of the posteriors of δ and σ_κ^2. In accordance with the BA-CBD model, the drift δ increases, since the steep estimates of the last decade are omitted. On the other hand, the main difference in the variance σ_κ^2 is the development of a heavier right tail and the increase of the posterior variance, rather than a shift in the location of the posterior. The greatest difference between the two calibrations of the BA-LC model occur for the age effects series, $\beta_x^{(2)}$. The underlying posterior distributions are tilted and generally become more volatile from one age to another, though the width of the credibility bands are generally consistent across the two different implementations.

The short term dynamics of the BA-LC for the truncated data-set are assessed in Figure 6.25, where the posterior densities of the autoregression parameters α_4, α_5 and α_6 are compared to those produced by the calibration of the model using the full EW data. In contrast to the BA-CBD model results, parameters α_4 and α_6 are centred around almost identical values for the two different implementations. The posterior uncertainty of α_4 and α_6 increases for this case too under the truncated estimation. Parameter α_5 is negligibly shifted downwards, implying again less strong cohort impact. However, the absolute difference of the location of the two posteriors cannot cause great differences in the projected residuals states. Figure 6.25 also includes the plot of the mean posterior estimate for the covariance matrix V_Z. The plot may be directly compared to the relevant estimate of the BA-LC under the full EW data given in Figure 6.7. Evidently, the two matrices include values within an almost identical range and maintain the exact same patterns.

Having examined the posterior distributions of the main components of the BA-models by using a subset of the originally used EW data-set, we conclude that the parameters of the models maintain their interpretation and form in both cases. The only noticeable difference that may be mentioned is the weaker cohort impact under the restricted data-set, which is more prominent for the BA-CBD model. However, the main characteristics of the long and short term dynamics of the models are consistent in both cases.

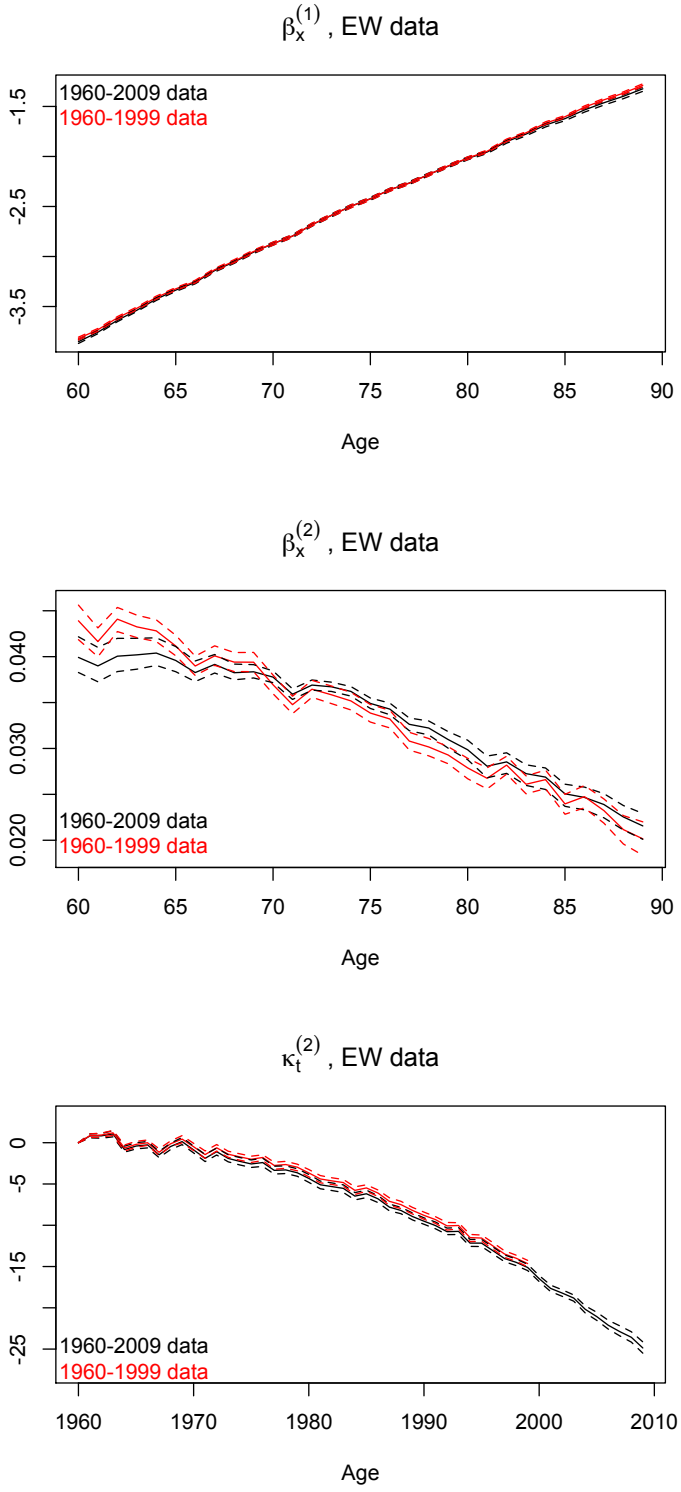

Figure 6.24: Posterior distributions of age effects series, $\beta_x^{(1)}$ and $\beta_x^{(2)}$, and of the period effects series, $\kappa_t^{(2)}$, of the BA-LC model for the EW data, using data 1960-2009 and 1960-1999, in black and red lines respectively. Solid lines indicate mean estimates and dashed lines show the 95% credible intervals.

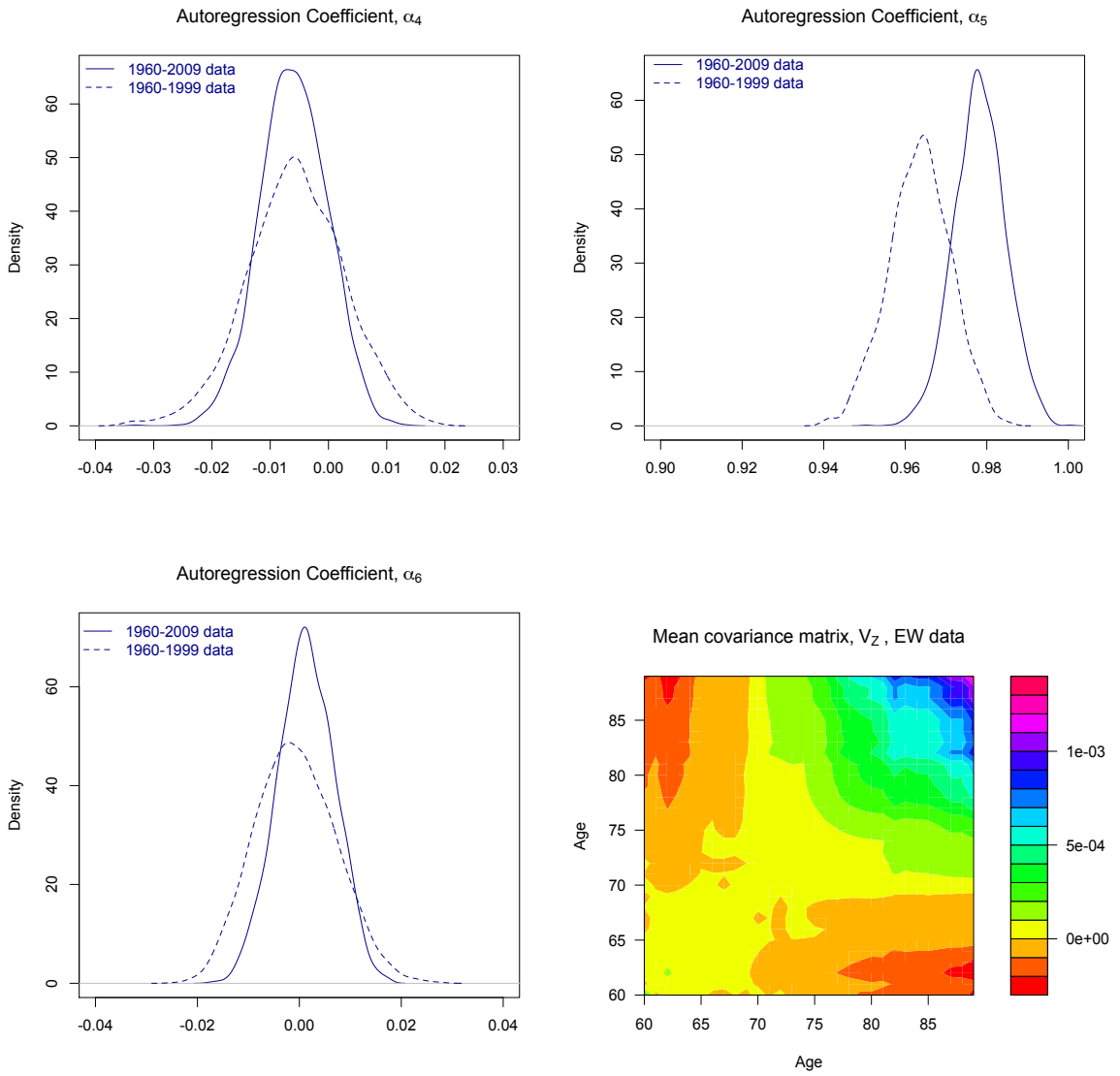

Figure 6.25: Posterior densities of the autoregression coefficients, α_4, α_5 and α_6, of the BA-LC model using the full and truncated EW data, and mean posterior estimate of the covariance matrix V_Z for the truncated EW data.

Based on these dynamics the BA-models are projected using estimates up to year 1999. Figures 6.26 and 6.27 show the realised mortality or death rates on the appropriate scale, respectively, along with in- and out-of-sample forecasts of the BA-models. The in-sample forecast is for 30 years ahead starting from 2000 and the out-of-sample projection is of the following 20 years starting from 2010. Each plot also includes in-sample projections for the other models of the family as appropriate. Hence, we can first examine the difference in the predicted mortality trend under the BA-models if the last decade of the data-set is excluded and how this is compared to the full forecasts of the models. Second, we can assess which one of the models performed better in estimating the mortality patterns developed in the last 10 years.

The plots give rise to for several comments. The increased estimates and posterior distributions of the drifts for years 1960-1999 are associated with the greater forecasted rates of the LC, CBD and their BA-forms compared to the mortality realisation. Model M7 appears to be following an upwards shift as soon as projected cohort factors enter the projections for both ages 75 and 85. Notably, for age 75 the mortality dynamics are forecasted very accurately, but when projected cohort effects are incorporated the forecasts get erratic. That upwards trend must also be connected to greater drift estimates.

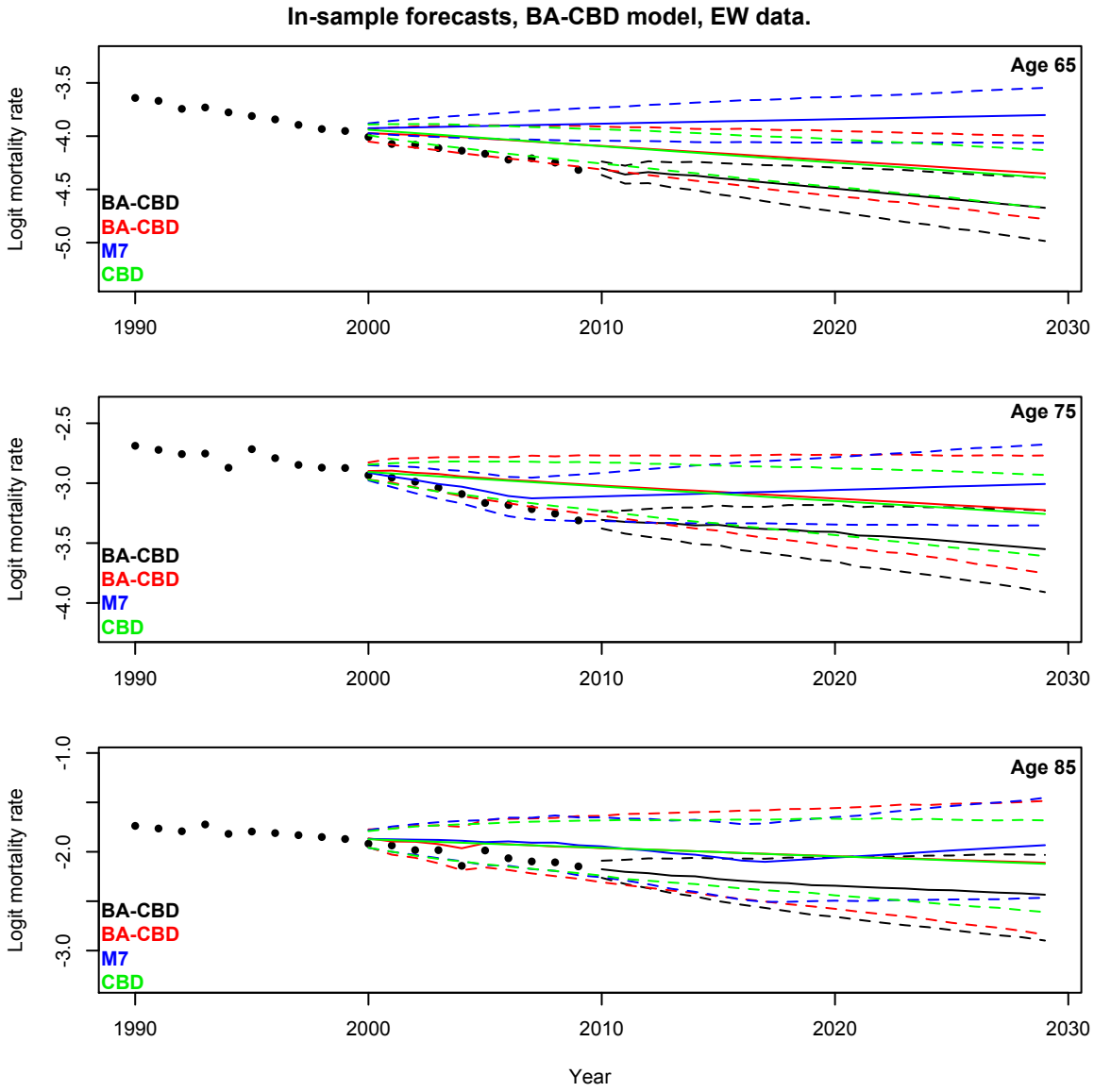

Figure 6.26: In sample forecasts for the BA-CBD, CBD and M7 models using the truncated EW data, compared to the forecast of the BA-CBD model under the full EW data. Solid lines indicate the mean projection and the dashed lines show the 95% confidence bands.

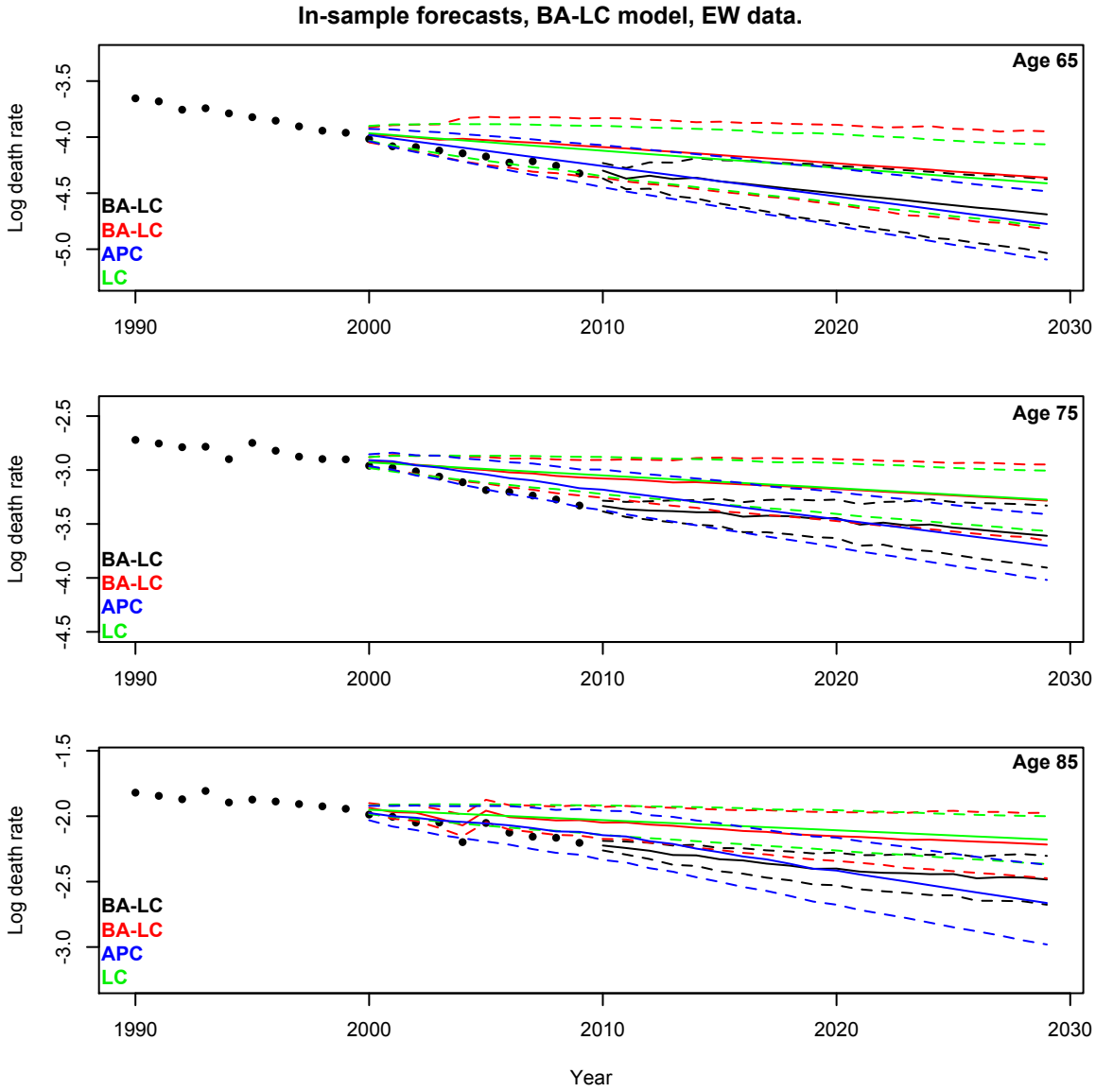

Figure 6.27: In sample forecasts for the BA-LC, LC and APC models using the truncated EW data, compared to the forecast of the BA-LC model under the full EW data. Solid lines indicate the mean projection and the dashed lines show the 95% confidence bands.

The confidence bands of the BA-CBD model get distinctively wider than those of the original CBD model, especially as the examined age increases. On the other hand, the joint residuals projection only contributes a small amount of additional variability to the forecasts of the LC model. The APC model is again the one which produces the steepest forecasts, as was the case with the projections of Section 6.4 for the full data-set. Amazingly, the forecasts of the APC model are those that are closer to the realised mortality trend of the decade 2000-2009 than any other of the stochastic mortality models examined. If this would also be the case for the following, say, 20 years mortality indices would dramatically fall and measures such as life-expectancies would rise significantly, as shown in Section 1.4.3. A final comment regarding the

information of the data that is incorporated in the forecasts. It can be seen in the above plots for age 85 that both the BA-models capture the jump of the cohort of 1919 in 2004. This is also strongly connected to the comment about the effect of the residuals process in the projections in Section 6.4. The cohort enhanced models, APC and M7, do not capture that effect, solely because this information has been removed; recall that the cohorts of 1918, 1919 and 1920 were omitted for the EW data in Chapter 1. If these were included in the estimation of the ARIMA model for the cohort factors the jump would be reproduced therein too. Hence, the BA-models capture and incorporate by construction the full patterns and information contained in the modelled data-set.

6.7 Summary

This Chapter aided in conducting an in depth qualitative and quantitative comparison of the stochastic mortality models used in the Thesis for the EW and Canada populations.

The fitted rates of the BA-models almost match the historical mortality rates. Hence, the behaviour of the standardised mortality residuals is consistent with underlying modelling assumptions, as shown in Section 6.1. The posterior variability of the BA-models dynamics acts sort of inversely for the two examined populations. The long-term parameters are persistently subject to greater uncertainty for EW, but the opposite is true for the local mortality factors as described through matrices $\mathcal{A}$ and V_Z. This effect is also connected to the stronger structures observed within the joint posteriors for the elements of matrix $\mathcal{A}$ for the Canada population in Sections 4.3.2 and 5.3.2. The BA-CBD model is seen to perform the best of those employed in terms of the DIC in all cases.

In terms of forecasting properties, the BA-models incorporate by construction a great amount of information within the data. The confidence levels get in some cases noticeably wider and observed patterns are being reproduced. Accordingly, the simulated distributions of the survivor indices develop greater spread, which is inherited by the calculations of related mortality metrics. The projected mortality improvements over the whole projection period showed that the models that incorporate cohort effects produce more extreme values for the initial cohorts of the population. As soon as those cohorts exit the age window of the forecasted mortality table, improvements fall almost by half, which is of course due to the loss of that information. On the other hand, the LC and CBD models that do not take into account the cohort information result in balanced improvement levels over the whole projection period.

By restricting the original EW data-set by 10 years the BA-models are refitted and the results were used to examine their robustness properties. The interpretation of their parameters remains consistent and the various types of dynamics are very similar, allowing for differences due to the qualitative characteristics of the 2000-'09 decade's period effects. The results of the in-sample forecasts follow from the differences in the models dynamics as commented in Section 6.6. Most interestingly, the APC model which produces in some sense the most erratic forecasts is the one that is closer to the realised rates of the decade 2000-'09.

Epilogue

Throughout this Thesis extended versions of the well-known LC and CBD stochastic mortality models were developed. Based on the statistical properties of their estimated residuals matrices under the Poisson model for deaths, an additional multivariate stochastic process was employed to incorporate the full information of the examined data-sets. By relaxing the conditionally independent residuals assumption of the Poisson model, we introduce serial dependence within the cells of the population in the time dimension. Additional correlation in the age dimension may also be incorporated within the construction of the residuals model. Practical applications showed that such a structure is necessary when using the LC model as a benchmark, mainly due to its single stochastic component. In that case, the Bayesian approach is beneficial due to semi-conjugate properties of the IW distribution within the model. The residuals augmentations method, under the adopted Bayesian paradigm resulted in naming the suggested method Bayesian Augmentation. Based on several qualitative and quantitative criteria, the impact of the residuals process was analysed in isolation, by focusing on the results of the BA-models and of the basic versions of the mortality models. The identical comparison was also conducted between the BA-models and cohort-enhanced versions of the LC and CBD models, namely the APC model and model M7.

Methods

The general framework of latent states augmentation arises naturally under several different problems. The Bayesian approach seems to be an efficient implementation choice, especially in mixture cases where the posteriors are analytically intractable. For example, Johannes and Polson (2009) develop in some cases similar algorithms to those used herein for standard Econometric problems. In the present case, the hybrid MCMC scheme utilised prior information in a way such that the posteriors are tractable or close to tractable, so that an efficient approximation may be identified. In that sense, the parameters of the long-term mortality dynamics resulted from Gibbs sampling, whereas MH steps were used for the majority of the involved latent states and the parameters of the autoregression matrix, $\mathcal{A}$, of the models. The BA-models

may be classified as Generalised (Non/)Linear Mixed Effects models. The Linear case is reserved for the BA-CBD model, whereas the BA-LC model belongs to the Non-linear class due to its age-period interaction terms. The generic issue of Bayesian inference for such models has been presented by Wakefield (2013), where the modelling techniques and implementations of the Thesis have been sketched out.

The suggested augmentation essentially imposes structured random effects, which in the mortality modelling context might be interpreted as the local mortality effects of the modelled population. For example, under our parameterisation, the residuals process carries the impact of the cohort effect along with components of adjacent cohorts. Although there are cases where the influence of the latter cohorts is negligible, these effects may be considered local in the sense of being internal for a population. As the additional residuals model is stationary, the long-term, persistent mortality characteristics are still determined by the random period factor(s), respectively for each case. Similar random effects Bayesian stochastic mortality models may also found in Cairns (2013), although not structured and without a clear natural interpretation of what they represent. From another point of view, D'Amato et al. (2012) consider the correlation structure across time within the residuals of the Lee-Carter model, which is model by an autoregressive process of order determined according to some information criterion, *e.g.* AIC. In a multi-population framework, Yang and Wang (2013) model the first differenced residuals of the Lee-Carter model as zero mean Normal variables independently for each age. Subsequently, they introduce further age-specific correlation across the various populations they examine. In contrast to the above three approaches, the present work suggests a framework which allows for modelling serial and contemporary correlation simultaneously. The methods for estimating jointly the latent states and the parameters of the BA-models follow closely that employed by Czado et al. (2005) in a single population framework, and Cairns et al. (2011b) under a two populations approach.

Conclusions

The results of the Thesis show that the residuals augmentation does not alter qualitatively the long-term dynamics of the mortality modelling bases. In contrast, the shape of the interaction age effects, where these exist, may be be slightly distorted. The slope parameters $\kappa_t^{(2)}$ of the BA-CBD model and the $\beta_x^{(2)}$ age effects of the BA-LC model develop the highest amount of posterior variability. The latent residuals compensate for the missing structures of the benchmark models, and the specification of the model feeds in the projection of the historical mortality characteristics of the population. Once the initial cohorts of the population exit the age limit of the projected mortality table, the produced forecasts are smooth. The same effect

is also observed for the APC and model M7, where projections become smooth once simulated, instead of estimated, cohort effects are incorporated. The results of that behaviour were seen in Section 6.5, where the forecasted improvements of the BA-models, the APC and of model M7 behave distinctly for the two periods. On the other hand, the range of the improvements under the LC and CBD models are stable throughout the projection horizon.

The samples from the posterior distributions of the BA-models provide information which would have been otherwise inadmissible. The parameters of the long-term dynamics have higher posterior variability for the EW than for the Canada data, but the opposite is true for the parameters of the local mortality dynamics, *e.g.* the elements of the autoregression matrix A. The joint posterior of the drift vector, $\boldsymbol{\delta}$, of the BA-CBD model develops significant correlation, but this is an effect of the joint residuals model simulation, since that dependency was not observed with the B-CBD model of Section 2.4.1. Within the residuals model, the diffusion coefficient α_2 seems to act inversely to the cohort-type factor α_3 for both BA-forms and data-sets, although the effects are consistently more intense for Canada. Structured relationships are also observed within the elements of the triple $(\alpha_4, \alpha_5, \alpha_6)$, where again the results are more prominent for the Canada data. Such information would not be available naturally and in such detail under a frequentist approach, where bootstrap methods would need to be used to assess parameter uncertainty. If one considers the combined effort of fitting and assessing parameter uncertainty of a statistical model, becomes obvious that the Bayesian approach is more informative, and sometimes also favoured, as the complexity of the model increases.

The fitted values of the BA-models almost match the historical data and projections are initiated almost from the observed rates. The performance of the standardised residuals is in broad agreement with the modelling assumptions, as also are the estimated latent residual states, $\mathcal{R}(x,t)$. On the one side, the residuals model perform adequately in supplementing qualitatively the employed mortality modelling structures. On the other hand, both suggested models have drawbacks. The BA-LC model still lacks the essential multi-factor structure and the BA-CBD lacks an age-level set of parameters. Given the results of the t-tests over the dimensions of the posterior distribution of the standardised residuals matrices of the BA-CBD model, it is likely that the most appropriate benchmark could be in the form of:

$$\operatorname{logit}\big(q(x,t)\big) = \alpha_x + \kappa_t^{(1)} + \kappa_t^{(2)}\big(x - \bar{x}\big).$$

The increased variability of the projections under the BA-models is generally inherited to mortality metrics and indices, but the relative differences are small. After fitting the BA-models to a subset of the original full EW data, the interpretation of their

parameters is kept consistent, verifying the robustness of the suggested extensions. The most notable difference could be observed in the weakening of the cohort-type factor α_5 for both BA-forms. The results of the in-sample forecasts show that the majority of the models fail to capture the downwards mortality trend in EW during the 2000-'09 decade. An exception is provided by the APC model which appears the only one following the right pattern. Considering that forecasts under the APC model in Sections 1.4.2 and 6.4 appeared as the most erratic compared to the other stochastic mortality models, one understands that model risk within the mortality modelling context should be taken into account very seriously.

Appendix A

Markov Chains

This Appendix provides the mathematical definitions of the properties a Markov chain oughts to satisfy in order for it to possess a stationary distribution. A concept required for the following definitions is that of T_{yy}, the time of the first return of the chain $Y^{(n)}$ to some state $y \in \mathcal{S}$, i.e. $T_{yy} = \min\{k \geq 1 : Y^{(k)} = y | Y^{(0)} = y\}$.

Definition A.1 *The Markov chain $Y^{(n)}$ is called irreducible if for all states $x, y \in \mathcal{S}$, there exists $n \geq 1$ such that $\mathbb{P}^n(x, y) > 0$.*

Definition A.2 *The Markov chain $Y^{(n)}$ is called aperiodic if for all states $y \in \mathcal{S}$, the largest integer t, such that all the times at which the chain returns to y are multiplies of t, is equal to 1.*

Definition A.3 *An irreducible Markov chain $Y^{(n)}$ is called recurrent if for all states $y \in \mathcal{S}$, $\mathbb{P}(T_{yy} < \infty) = 1$.*

Definition A.4 *An irreducible and recurrent Markov chain $Y^{(n)}$ is called positive recurrent if for all states $y \in \mathcal{S}$, $\mathbb{E}(T_{yy}) < \infty$. Equivalently, $Y^{(n)}$ is positive recurrent if and only if there exists a stationary probability distribution $\pi(\cdot)$ for $Y^{(n)}$, that is if and only if there exists $\pi(\cdot)$ such that*

$$\sum_y \pi(y)\mathbb{P}(y, x) = \pi(x), \text{ for all } x \in \mathcal{S}.$$

Definition A.5 *The Markov chain $Y^{(n)}$ is called time reversible if it is positive recurrent with invariant distribution $\pi(\cdot)$ and*

$$\pi(x)\mathbb{P}(x, y) = \pi(y)\mathbb{P}(y, x), \text{ for all } x, y \in \mathcal{S}.$$

The first four of the above definitions were used to establish the result of the ergodicity theorem stated in the end of Section 2.2.1. The fifth is the detailed balance equations, required for the construction of the MCMC algorithms.

Appendix B

Statistical Tests

This Appendix provides the details of the statistical tests employed in Sections 3.2 and 6.1.

For univariate cases the observations are noted $\boldsymbol{x} = (x_1, \ldots, x_n)$, where n is the sample size. The sample mean, the unbiased estimator of the population variance and the estimated standard error of the population mean are:

$$\begin{aligned} \bar{x} &= \frac{\sum_i x_i}{n} \\ s^2 &= \frac{1}{n-1}\sum_i (x_i - \bar{x})^2 \\ s_{\bar{x}} &= \frac{s}{\sqrt{n}}. \end{aligned}$$

The sample skewness and kurtosis are:

$$\begin{aligned} S &= \frac{1}{n}\sum_i (x_i - \bar{x})^3 \Big/ s^3 \\ K &= \frac{1}{n}\sum_i (x_i - \bar{x})^4 \Big/ s^4. \end{aligned}$$

For multivariate cases, the p-variate observations are noted $\mathbf{X} = (X_1, \ldots, X_n)$, and n as before. The sample mean and the unbiased estimator of the population covariance matrix are:

$$\begin{aligned} \overline{X} &= \frac{1}{n}\sum_i X_i \\ W &= \frac{1}{n-1}\sum_i \big(X_i - \overline{X}\big)\big(X_i - \overline{X}\big)'. \end{aligned}$$

The confidence level for each of the following cases is denoted by a.

Runs test for independence

Definition

A run is defined to be a succession of one or more identical symbols which are followed and preceded by a different symbol or no symbol at all. Clues to lack of randomness are provided by any tendency of the symbols to exhibit a definite pattern in the sequence.

Assumptions

1) The data may be divided into two types, n_1 of type 1 and n_2 of type 2, so that $n_1 + n_2 = n$.
2) The total number of runs, r, consists of r_1 runs of type 1 and r_2 runs of type 2, so that $r_1 + r_2 = r$.

Null vs Alternative hypothesis

H_0: Each element in the sequence is independently drawn from the same distribution.
H_1: The elements in the sequence are not independent.

Computation

Inference is based on the random variable, R, denoting the number of runs when the hypothesis of randomness is true, whose exact distribution is available, *e.g.* see Gibbons and Chakraborti (2011). If $n_1, n_2 > 10$ and $n \to \infty$ so that $\lambda = n_1/n$ and $1 - \lambda = n_2/n$ remains constant then:

$$\begin{aligned} \lim_{n\to\infty} \mathbb{E}\big(R/n\big) &= 2\lambda(1-\lambda) \\ \lim_{n\to\infty} Var\big(R/\sqrt{n}\big) &= n\lambda^2(1-\lambda)^2. \end{aligned}$$

It can be shown that R is normally distributed, the standardised test statistic is:

$$S = \frac{R - 2n\lambda(1-\lambda)}{2\sqrt{n}\lambda(1-\lambda)},$$

and for the two-sided alternative employed herein, the null hypothesis can be rejected when $|S| > z_{a/2}$, where $z_{a/2}$ is the $(1 - a/2)$th quantile point of the standard normal distribution.

The test is available in R within the 'randtest' package (Caeiro and Mateus, 2014). The routine offers the choice of *exact* or *asymptotic* calculation, which was employed as appropriate.

t-test for mean

Assumptions

1) The sample is randomly selected from the population it represents.

2) The data are normally distributed.

Null vs Alternative hypothesis

H_0: $\mu = \mu_0$

H_1: $\mu \neq \mu_0$.

Computation

The t-statistic:

$$t = \frac{\bar{x} - \mu}{s_{\bar{x}}},$$

is distributed according to the t_{n-1} distribution, and for the two-sided alternative employed herein, the null hypothesis can be rejected when $|t| > t_{n-1,a/2}$, where $t_{n-1,a/2}$ is the $(1 - a/2)$th quantile point of the t distribution with $n - 1$ degrees of freedom.

Hotelling's T^2-test for multivariate mean

Assumptions

1) The sample consists of independent observations.

2) The data follow the multivariate normal distribution.

Null vs Alternative hypothesis

H_0: $\boldsymbol{\mu} = \boldsymbol{\mu}_0$

H_1: $\boldsymbol{\mu} \neq \boldsymbol{\mu}_0$.

Computation

The T^2-statistic:

$$T^2 = n\big(\overline{X} - \boldsymbol{\mu}\big)' W^{-1} \big(\overline{X} - \boldsymbol{\mu}\big),$$

has an exact admissible distribution, namely the Hotelling's T^2 distribution. Due to the estimate used for the covariance matrix of the population, it is also approximately χ^2_p distributed. For the following transformation holds

$$F = \frac{n - p}{p(n-1)} T^2 \sim F_{p,n-p},$$

where $F_{p,n-p}$ is Fisher's F distribution with n and $n - p$ degrees of freedom and the null hypothesis can be rejected when $F > F_{p,n-p,a}$, where $F_{p,n-p,a}$ is the critical value of the F distribution at significance level a. Rejection of the test implies that at least one of the p means is not equal to its hypothesised value.

χ^2-test for variance

Assumptions

1) The sample is randomly selected from the population it represents.

2) The data are normally distributed.

Null vs Alternative hypothesis

H_0: $\sigma^2 = \sigma_0^2$

H_1: $\sigma^2 \neq \sigma_0^2$.

Computation

The χ^2-statistic:

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2},$$

is distributed according to χ^2_{n-1} distribution, and for the two-sided alternative employed herein, the null hypothesis can be rejected when $\chi^2 \notin (\chi^2_{n-1,a/2}, \chi^2_{n-1,1-a/2})$, where $\chi^2_{n-1,a/2}$ and $\chi^2_{n-1,1-a/2}$ are the $a/2$th and $(1-a/2)$th quantile points of the χ^2 with $n-1$ degrees of freedom.

Jarque-Bera (JB) test for normality

Assumptions

1) The sample is randomly selected from the population it represents.

2) The sample consists of independent observations.

Null vs Alternative hypothesis

H_0: The data are normally distributed.

H_1: The data are not normally distributed.

Computation

The JB-statistic:

$$JB = \frac{n}{6}\Big(S^2 + \frac{1}{4}\big(K-3\big)^2\Big),$$

is distributed according to χ^2_2 distribution, and the null hypothesis can be rejected when $JB > \chi^2_{2,1-a}$, where $\chi^2_{2,1-a}$ is the $(1-a)$th quantile points of the χ^2 with 2 degrees of freedom.

Bibliography

Anderson, T. W. (1984). ***"An Introduction to Multivariate Statistical Analysis"***. Wiley.

Andrieu, C. and Thoms, J. (2008). **"A tutorial on adaptive MCMC"**. *Statistics and Computing*, 18(4): 343–373.

Atchade, Y., Fort, G., Moulines, E., and Priouret, P. (2009). **"Adaptive markov chain monte carlo: theory and methods"**. *Preprint*.

Beard, R. E. (1971). **"Some aspects of theories of mortality, cause of death analysis, forecasting and stochastic processes."**. *Biological aspects of demography*, pages 57–68.

Besag, J. (1974). **"Spatial interaction and the statistical analysis of lattice systems"**. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 192–236.

Biffis, E. (2005). **"Affine processes for dynamic mortality and actuarial valuations"**. *Insurance: mathematics and economics*, 37: 443–468.

Biffis, E. and Blake, D. (2009). **"Mortality Linked Securities and Derivatives"**. *Cass Business School Pensions*.

Blake, D., Cairns, A. J., and Dowd, K. (2006). **"Living with mortality: Longevity bonds and other mortality-linked securities"**. *British Actuarial Journal*, 12: 153–197.

Booth, H., Hyndman, R. J., Tickle, L., De Jong, P., et al. (2006). **"Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions'**. Technical report, Monash University, Department of Econometrics and Business Statistics.

Booth, H., Maindonald, J., and Smith, L. (2002). **"Applying Lee–Carter under conditions of variable mortality decline"**. *Population Studies*, 56: 325–336.

Booth, H. and Tickle, L. (2008). **"Mortality modelling and forecasting: A review of methods"**. *Annals of Actuarial Science*, 3: 3–43.

Box, G. E., Jenkins, G. M., and Reinsel, G. C. (2013). ***"Time series analysis: forecasting and control"***. Wiley.

Brooks, S. P. and Roberts, G. O. (1998). **"Assessing convergence of Markov chain Monte Carlo algorithms"**. *Statistics and Computing*, 8: 319–335.

Brouhns, N., Denuit, M., and Van Keilegom, I. (2005). **"Bootstrapping the Poisson log-bilinear model for mortality forecasting"**. *Scandinavian Actuarial Journal*, 2005: 212–224.

Brouhns, N., Denuit, M., and Vermunt, J. K. (2002). **"A Poisson log–bilinear regression approach to the construction of projected life–tables"**. *Insurance: Mathematics and Economics*, 31: 373–393.

Caeiro, F. and Mateus, A. (2014). ***"randtests: Testing randomness in R"***. R package, http://CRAN.R-project.org/package=randtests, 0.3 edition.

Cairns, A. J. G. (2000). **"A discussion of parameter and model uncertainty in insurance"**. *Insurance: Mathematics and Economics*, 27: 313–330.

Cairns, A. J. G., Blake, D., and Dowd, K. (2006a). **"A Two–Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration"**. *Journal of Risk and Insurance*, 73: 687–718.

Cairns, A. J. G., Blake, D., and Dowd, K. (2006b). **"Pricing death: Frameworks for the valuation and securitisation of mortality risk"**. *ASTIN Bulletin*, 36: 79–120.

Cairns, A. J. G., Blake, D., and Dowd, K. (2008). **"Modelling and management of mortality risk: a review"**. *Scandinavian Actuarial Journal*, 2008: 79–113.

Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., and Khalaf-Allah, M. (2011a). **"Mortality density forecasts: An analysis of six stochastic mortality models"**. *Insurance: Mathematics and Economics*, 48: 355–367.

Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A., and Balevich, I. (2007). **"A quantitative comparison of stochastic mortality models using data from England and Wales and the United States"**. *Pre–print*.

Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A., and Balevich, I. (2009). **"A quantitative comparison of stochastic mortality**

models using data from England and Wales and the United States". *North American Actuarial Journal*, 13: 1–35.

Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., and Khalaf-Allah, M. (2011b). **"Bayesian stochastic mortality modelling for two populations"**. *ASTIN Bulletin*, 41: 29–59.

Cairns, A. J. G., Blake, D., Dowd, K., and Kessler, A. (2014). **"Phantoms never die: Living with unreliable mortality data"**. *Working Paper, Heriot-Watt University.*

Cairns, G. L. (2013). ***"A Bayesian approach to modelling mortality, with applications to insurance"***. PhD thesis, University of Glasgow.

Cannon, E. (2009). **"Estimation and pricing with the Cairns-Blake-Dowd model of mortality"**. *University of Verona, Department of Economics*, Working Paper 65.

Carlin, B. P. and Louis, T. A. (2000). ***"Bayes and empirical Bayes methods for data analysis"***. CRC Press.

Casella, G. (1985). **"An introduction to empirical Bayes data analysis"**. *The American Statistician*, 39(2): 83–87.

Celeux, G., Forbes, F., Robert, C. P., Titterington, D. M., et al. (2006). **"Deviance information criteria for missing data models"**. *Bayesian Analysis*, 1: 651–673.

Chan, W.-S., Li, J. S.-H., and Cheung, S.-H. (2008). **"Testing deterministic versus stochastic trends in the Lee-Carter mortality indexes and its implications for projecting mortality improvements at advanced ages"**. *Living to 100, SOA*.

Chan, W.-S., Li, J. S.-H., and Li, J. (2014). **"The CBD mortality indexes: modeling and applications"**. *North American Actuarial Journal*, 18: 38–58.

Chiang, C. L. (1984). ***"The life table and its applications"***. Malabar (FL), Robert E. Krieger Publishing 1984.

CMI (2007a). **"Stochastic projection methodologies: Further progress and P–Spline model features, example results and implications"**. Working Paper 20.

CMI (2007b). **"Stochastic projection methodologies: Lee–Carter model features, example results and implications"**. Working Paper 25.

Coale, A. J. and Kisker, E. E. (1990). **"Defects in data on old–age mortality in the United States: New procedures for calculating mortality schedules and life tables at the highest ages."**. *Asian and Pacific Population Forum*, 4: 1–31.

Coughlan, G., Epstein, D., Alen, O., Sinha, A., Hevia-Portocarrero, J., Gingrich, E., Khalaf-Allah, M., and Joseph, P. (2007). **"LifeMetrics: A toolkit for measuring and managing longevity and mortality risks"**. Technical report, Pension Advisory Group JP Morgan Chase.

Cowles, M. K. and Carlin, B. P. (1996). **"Markov chain Monte Carlo convergence diagnostics: a comparative review"**. *Journal of the American Statistical Association*, 91: 883–904.

Currie, I. D. (2006). **"Smoothing and forecasting mortality rates with P-splines"**. In *Presentation to the IFoA*.

Currie, I. D., Durban, M., and Eilers, P. (2004). **"Smoothing and forecasting mortality rates"**. *Statistical modelling*, 4: 279–298.

Czado, C., Delwarde, A., and Denuit, M. (2005). **"Bayesian Poisson log–bilinear mortality projections"**. *Insurance: Mathematics and Economics*, 36: 260–284.

Dahl, M. (2004). **"Stochastic mortality in life insurance: Market reserves and mortality-linked insurance contracts"**. *Insurance: mathematics and economics*, 35: 113–136.

D'Amato, V., Haberman, S., Piscopo, G., and Russolillo, M. (2012). **"Modelling dependent data for longevity projections"**. *Insurance: Mathematics and Economics*, 51: 694–701.

De Jong, P. and Tickle, L. (2006). **"Extending Lee–Carter mortality forecasting"**. *Mathematical Population Studies*, 13: 1–18.

Dellaportas, P., Smith, A. F. M., and Stavropoulos, P. (2001). **"Bayesian analysis of mortality data"**. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 164: 275–291.

Delwarde, A., Denuit, M., and Eilers, P. (2007). **"Smoothing the Lee–Carter and Poisson log–bilinear models for mortality forecasting: A penalized log–likelihood approach"**. *Statistical Modelling*, 7: 29–48.

Dickson, D. C., Hardy, M. R., and Waters, H. R. (2013). ***"Actuarial mathematics for life contingent risks"***. Cambridge University Press.

Dowd, K., Cairns, A. J. G., Blake, D., Coughlan, G. D., Epstein, D., and Khalaf-Allah, M. (2010a). **"Backtesting Stochastic Mortality Models: An Ex–Post Evaluation of Multi–period Ahead Density Forecasts"**. *North American Actuarial Journal*, 14: 281–298.

Dowd, K., Cairns, A. J. G., Blake, D., Coughlan, G. D., Epstein, D., and Khalaf-Allah, M. (2010b). **"Evaluating the goodness of fit of stochastic mortality models"**. *Insurance: Mathematics and Economics*, 47: 255–265.

Eilers, P. H. and Marx, B. D. (1996). **" Flexible smoothing with B-splines and penalties"**. *Statistical science*, pages 89–102.

Forfar, D. O., McCutcheon, J. J., and Wilkie, A. D. (1988). **"On graduation by mathematical formula"**. *Journal of the Institute of Actuaries*, 115: 1–149.

Gamerman, D. and Lopes, H. F. (2006). ***"Markov chain Monte Carlo: Stochastic simulation for Bayesian inference"***, volume 68. CRC Press.

Gelfand, A. E. and Smith, A. F. (1990). **"Sampling-based approaches to calculating marginal densities"**. *Journal of the American statistical association*, 85: 398–409.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003). ***"Bayesian data analysis"***. CRC press.

Geweke, J. (1992). **"Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments"**. In *Bayesian Statistics*, pages 169–193. Oxford University Press.

Gibbons, J. D. and Chakraborti, S. (2011). ***"Nonparametric statistical inference"***. Springer.

Haberman, S. and Renshaw, A. (2011). **"A comparative study of parametric mortality projection models"**. *Insurance: Mathematics and Economics*, 48: 35–55.

Hastings, W. K. (1970). **"Monte Carlo sampling methods using Markov chains and their applications"**. *Biometrika*, 57: 97–109.

Heligman, L. and Pollard, J. H. (1980). **"The age pattern of mortality"**. *Journal of the Institute of Actuaries*, 107: 49–80.

Hobert, J. P. and Casella, G. (1996). **"The effect of improper priors on Gibbs sampling in hierarchical linear mixed models"**. *Journal of the American Statistical Association*, 91: 1461–1473.

Hollman, F. W., Mulder, T. J., and Kallan, J. E. (2000). **"Methodology and assumptions for the population projections of the United States: 1999 to 2100"**. *U.S. Bureau of the Census, Population Division: Working Paper 38*.

Hotelling, H. (1992). ***"The generalization of Student's ratio"***. Springer.

Hyndman, R. J., Booth, H., Tickle, L., and Maindonald, J. (2014). ***"Demography: Forecasting mortality, fertility, migration and population data"***, R package 1.17 edition.

Hyndman, R. J. and Ullah, M. S. (2007). **"Robust forecasting of mortality and fertility rates: A functional data approach"**. *Computational Statistics & Data Analysis*, 51: 4942 – 4956.

Jarque, C. M. and Bera, A. K. (1987). **"A test for normality of observations and regression residuals"**. *International Statistical Review/Revue Internationale de Statistique*, 55: 163–172.

Johannes, M. and Polson, N. (2009). **"Markov Chain Monte Carlo"**. In *Handbook of Financial Time Series*, pages 1001–1013. Springer.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1995). ***"Continuous univariate distributions "***, volume 2. Wiley.

Koissi, M.-C. and Shapiro, A. F. (2006). **"Fuzzy formulation of the Lee–Carter model for mortality forecasting"**. *Insurance: Mathematics and Economics*, 39: 287–309.

Lee, R. D. and Carter, L. R. (1992). **"Modeling and forecasting US mortality"**. *Journal of the American statistical association*, 87: 659–671.

Lee, R. D. and Miller, T. (2001). **"Evaluating the performance of the Lee–Carter method for forecasting mortality"**. *Demography*, 38: 537–549.

Li, J. S.-H. (2010). **"Pricing longevity risk with the parametric bootstrap: A maximum entropy approach"**. *Insurance: Mathematics and Economics*, 47: 176–186.

Li, J. S.-H., Chan, W.-S., and Cheung, S.-H. (2011). **"Structural changes in the Lee-Carter mortality indexes: detection and implications"**. *North American Actuarial Journal*, 15: 13–31.

Li, J. S.-H., Hardy, M., and Tan, K. S. (2009). **"Uncertainty in Mortality Forecasting'**. *ASTIN Bulletin*, 39: 137–164.

Li, J. S.-H., Hardy, M., and Tan, K. S. (2010). **"Developing Mortality Improvement Formulas: The Canadian Insured Lives Case Study"**. *North American Actuarial Journal*, 14: 381–399.

Lütkepohl, H. (2005). ***"New introduction to multiple time series analysis"***. Cambridge University Press.

Macdonald, A., Cairns, A., Gwilt, P., and Miller, K. (1998). **"An international comparison of recent trends in population mortality"**. *British Actuarial Journal*, 4: 3–141.

McNown, R. and Rogers, A. (1989). **"Forecasting mortality: A parameterized time series approach"**. *Demography*, 26: 645–660.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). **"Equation of state calculations by fast computing machines"**. *The journal of Chemical Physics*, 21: 1087–1092.

Milevsky, M. A. and Promislow, D. S. (2001). **"Mortality derivatives and the option to annuitise"**. *Insurance: Mathematics and Economics*, 29: 299–318.

ONS (2013a). **"National life-tables, England and Wales, 1980-82 to 2010-12"**. Technical report, Office for National Statistics.

ONS (2013b). **"Statistical Bulletin - Historic and Projected Mortality Data from the Period and Cohort Life Tables, 2012-based, UK, 1981-2062"**. Technical report, Office for National Statistics.

ONS (2014). **"Statistical Bulletin - Life Expectancy at Birth and at Age 65 by Local Areas in the United Kingdom, 2006-08 to 2010-12"**. Technical report, Office for National Statistics.

Osmond, C. (1985). **"Using age, period and cohort models to estimate future mortality rates"**. *International Journal of Epidemiology*, 14: 124–129.

Pedroza, C. (2006). **"A Bayesian forecasting model: predicting US male mortality"**. *Biostatistics*, 7: 530–550.

Perks, W. (1932). **"On some experiments in the graduation of mortality statistics"**. *Journal of the Institute of Actuaries*, 63: 12–57.

Pitacco, E., Denuit, M., Haberman, S., and Olivieri, A. (2009). ***"Modelling longevity dynamics for pensions and annuity business"***. Oxford University Press.

Plat, R. (2009). **"On stochastic mortality modeling"**. *Insurance: Mathematics and Economics*, 45: 393–404.

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). **"CODA: Convergence Diagnosis and Output Analysis for MCMC"**. *R News*, 6: 7–11.

Polson, N. G. (1996). **"Convergence of Markov Chain Monte Carlo Algorithms (with discussion)"**. In *Bayesian Statistics 5*, pages 297–323. Oxford University Press.

Renshaw, A. E. and Haberman, S. (2000). ***"Modelling for mortality reduction factors"***. City University, Department of Actuarial Science and Statistics.

Renshaw, A. E. and Haberman, S. (2003a). **"Lee–Carter mortality forecasting: a parallel generalized linear modelling approach for England and Wales mortality projections"**. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52: 119–137.

Renshaw, A. E. and Haberman, S. (2003b). **"Lee–Carter mortality forecasting with age–specific enhancement"**. *Insurance: Mathematics and Economics*, 33: 255–272.

Renshaw, A. E. and Haberman, S. (2006). **"A cohort–based extension to the Lee–Carter model for mortality reduction factors"**. *Insurance: Mathematics and Economics*, 38: 556–570.

Renshaw, A. E. and Haberman, S. (2008). **"On simulation–based approaches to risk measurement in mortality with specific reference to Poisson Lee–Carter modelling"**. *Insurance: Mathematics and Economics*, 42: 797–816.

Richards, S. J. (2008). **"Detecting year-of-birth mortality patterns with limited data"**. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171: 279–298.

Richards, S. J., Kirkby, J. G., and Currie, I. D. (2006). **"The importance of year of birth in two–dimensional mortality data"**. *British Actuarial Journal*, 12: 5–61.

Sheskin, D. J. (2003). ***"Handbook of parametric and nonparametric statistical procedures"***. CRC Press.

Smith, A. F. and Roberts, G. O. (1993). **"Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods"**. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 3–23.

Smith, B. J. (2007). **"BOA: An R Package for MCMC Output Convergence Assessment and Posterior Inference"**. *Journal of Statistical Software*, 21: 1–37.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). **"Bayesian measures of model complexity and fit"**. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64: 583–639.

Sweeting, P. J. (2011). **"A trend-change extension of the Cairns-Blake-Dowd model."**. *Annals of Actuarial Science*, 5: 143–162.

Tierney, L. (1994). **"Markov chains for exploring posterior distributions"**. *The Annals of Statistics*, pages 1701–1728.

Tsay, R. S. (2005). ***"Analysis of financial time series"***. Wiley.

Tuljapurkar, S., Li, N., and Boe, C. (2000). **"A universal pattern of mortality decline in the G7 countries"**. *Nature*, 405: 789–792.

Villasenor Alva, J. A. and Estrada, E. G. (2009). **"A generalization of Shapiro–Wilk's test for multivariate normality"**. *Communications in Statistics—Theory and Methods*, 38: 1870–1883.

Wakefield, J. (2013). ***"Bayesian and frequentist regression methods"***. Springer.

Willets, R. C. (2004). **"The Cohort Effect: Insights and Explanations"**. *British Actuarial Journal*, 10: 833–877.

Wilmoth, J. R. (1993). **"Computational methods for fitting and extrapolating the Lee–Carter model of mortality change"**. Technical report, Department of Demography, University of California, Berkeley.

Wilmoth, J. R., Andreev, K., Jdanov, D., and Glei, D. (2007). ***"Methods Protocol for the Human Mortality Database"***. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany), 5th edition.

Wong-Fupuy, C. and Haberman, S. (2004). **"Projecting mortality trends: recent developments in the United Kingdom and the United States"**. *North American Actuarial Journal*, 8: 56–83.

Yang, S. S. and Wang, C.-W. (2013). **"Pricing and securitization of multi-country longevity risk with mortality dependence"**. *Insurance: Mathematics and Economics*, 52: 157–169.